JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 1 (1971) pp. 1–26.

ON THE CATEGORY OF DIRECT SYSTEMS AND FUNCTORS ON GROUPS

Peter HILTON **
Battelle Institute, Geneva, Switzerland

Received 4 January 1970

§ 1. Introduction

Given a category $\mathfrak{C}$ and a directed set A, there is, of course, the well-known category $\mathfrak{C}^A$ of *direct systems in* $\mathfrak{C}$ *over* A. Moreover, if $\mathfrak{C}$ is cocomplete there is a colimit functor from $\mathfrak{C}^A$ to $\mathfrak{C}$. In certain contexts, however, it is desirable to be able simultaneously to consider direct systems in $\mathfrak{C}$ over variable directed sets. For example, every group is the union of its finitely-generated subgroups, but different groups plainly give rise in this way to direct systems of finitely-generated subgroups over different directed sets.

Our main objective in this paper is to describe a procedure for extending functors from a category of groups $\mathfrak{G}_0$ to the category $\mathfrak{G}_1$ of groups representable as colimits of direct systems of groups in * $\mathfrak{G}_0$. The procedure may be described as *extension by cocontinuity* and doubtless exists in abstract form in the literature; our emphasis here is on its relevance to the concrete category of groups. The original motivation for introducing this procedure for a category of groups was to introduce arbitrary Abelian coefficient groups into a general cohomology theory; this application will constitute the subject matter of a subsequent paper [6].

In § 2, we define the category $\mathfrak{C}^S$ of direct systems in $\mathfrak{C}$ over arbitrary directed sets. This category contains the category $\mathfrak{C}^A$ for each directed set A but is much more than their union. It turns out that an important technical notion in $\mathfrak{C}^S$ is that of a *fibre-map*; this notion specializes that of a fibred category in the sense of Gray and Grothendieck [4, 5] and enables us to define pull-backs in $\mathfrak{C}^S$ provided $\mathfrak{C}$ has pull-backs. Moreover if $\mathfrak{C}$ is cocomplete there is a direct limit functor $\underrightarrow{\lim}: \mathfrak{C}^S \to \mathfrak{C}$ which coincides on $\mathfrak{C}^A$ with the direct limit $\underrightarrow{\lim}^A$.

In § 3, we specialize to the case when $\mathfrak{C}$ is a category of groups. We use this

* Since we are exclusively concerned in this paper with index categories which are directed sets we speak of *direct limits* instead of colimits. Our results would generalize to *quasi-filtered* index categories (see [1, 2]).

** Present address: Cornell University, Ithaca, N.Y., USA.

specialization exclusively to establish the key lemma, which appears in the literature, that *direct limits commute with pull-backs* (Lemma 3.6). Thus any cocomplete category enjoying this property would suffice for our theory. We show that, with a mild hypothesis on $\mathfrak{G}_0$ and $\mathfrak{G}_1$, the category of fractions with respect to $\varinjlim$: $\mathfrak{G}_0^S \to \mathfrak{G}_1$ is equivalent to $\mathfrak{G}_1$; and this gives us the theorem on extending functors from $\mathfrak{G}_0$ to $\mathfrak{G}_1$ referred to above. The hypothesis is satisfied, for example, if $\mathfrak{G}_0$ is the category of finitely-generated Abelian groups, when $\mathfrak{G}_1$ is the category of Abelian groups; this is, essentially, the example with which our application in [6] is concerned. However, the hypothesis is violated if $\mathfrak{G}_0$ is the category of finitely-generated groups, when $\mathfrak{G}_1$ is the category of groups. In §4 we replace the original hypothesis by one satisfied in this case. The result we get is not quite as strong as in §3. Functors satisfying a certain rather obvious condition may, as before, be extended from $\mathfrak{G}_0$ to $\mathfrak{G}_1$; but we can only guarantee cocontinuity (see (3.19)) with respect to morphisms drawn from $\mathfrak{G}_0^A$ for some A.

In § 5 we prove a theorem on exactness in $\mathfrak{G}^S$ and direct limits and in §6 we make some remarks of an exploratory nature with respect to fibre-maps and their possible role in a hypothetical homotopy theory of directed sets. In §7 we describe in outline a related approach to the central problem of this paper, due to F.Ulmer.

The author had the benefit of many fruitful conversations with H.H.Storrer in developing the ideas deployed in this paper.

§ 2. Direct systems; fibre-maps; direct limits

Let $\mathfrak{C}$ be a category. Then the category of *direct systems in* $\mathfrak{C}$, written $\mathfrak{C}^S$, is defined as follows. An *object* of $\mathfrak{C}^S$ is a pair $(A; C.)$ where A is a directed set and $C. = \{C_\alpha; f_{\alpha\alpha'}\}$ is a direct system in $\mathfrak{C}$ over the directed set A. A *morphism* of $\mathfrak{C}^S$

$$\Phi : (A; C.) \to (B; D.) \quad (D. = \{D_\beta; g_{\beta\beta'}\})$$

is a pair $\Phi = (\rho, \varphi.)$, where $\rho : A \to B$ is a function which is (a) monotone and (b) cofinal; and $\varphi.$ associates with each $\alpha \in A$ a morphism $\varphi_\alpha : C_\alpha \to D_{\rho(\alpha)}$ such that, for $\alpha \leqslant \alpha'$, the diagram

$$(2.1) \qquad \begin{array}{ccc} C_\alpha & \xrightarrow{f_{\alpha\alpha'}} & C_{\alpha'} \\ \downarrow{\scriptstyle \varphi_\alpha} & & \downarrow{\scriptstyle \varphi_{\alpha'}} \\ D_{\rho(\alpha)} & \xrightarrow[g_{\rho(\alpha),\rho(\alpha')}]{} & D_{\rho(\alpha')} \end{array}$$

commutes.

It is evident that, with the obvious definition of composition of morphisms, $\mathfrak{C}^S$ is a category. Moreover, there is an evident embedding $P : \mathfrak{C} \to \mathfrak{C}^S$, whereby the object C is embedded as the direct system $\{C\}$ over a singleton set, so that we may introduce the notions of limit and colimit in the category $\mathfrak{C}^S$. It is also plain how a functor $F : \mathfrak{C} \to \mathfrak{T}$ induces a functor $F^S : \mathfrak{C}^S \to \mathfrak{T}^S$.

Definition 2.2. Let A, B be non-empty directed sets. A monotone function $\rho : A \to B$ is called a *fibre-map* if, given $\beta' \geqslant \rho(\alpha)$, there exists $\alpha' \geqslant \alpha$ with $\rho(\alpha') = \beta'$. A morphism $\Phi = (\rho, \varphi.)$ of $\mathfrak{C}^S$ is called a *fibre-map* if ρ is a fibre-map. (Notice that a fibre-map is automatically cofinal.)

We now introduce the *standard factorization* of a morphism of $\mathfrak{C}^S$. Thus, given $\Phi : (A; C.) \to (B; D.)$ as above, we define the diagram

$$(2.3) \qquad \begin{tikzcd} & (\bar{A}; \bar{C}.) \arrow[dl, bend right, "\bar{\Theta}"'] \arrow[d, "\bar{\Phi}"] \\ (A; C.) \arrow[ur, "\Theta"] \arrow[r, "\Phi"'] & (B; D.) \end{tikzcd}$$

as follows. $\bar{A} \subseteq A \times B$ consists of those pairs (α, β) with $\beta \geqslant \rho(\alpha)$; this is plainly a directed set under the order relation induced by the product order on $A \times B$. Further $\bar{C}_{\alpha\beta} = C_\alpha$ and the morphism $\bar{C}_{\alpha\beta} \to \bar{C}_{\alpha'\beta'}$, $\alpha \leqslant \alpha'$, $\beta \leqslant \beta'$, is just $f_{\alpha\alpha'}$. We next define $\Theta = (\sigma, 1.)$, where $\sigma(\alpha) = (\alpha, \rho(\alpha))$ and $1_\alpha = 1 : C_\alpha \to \bar{C}_{\alpha,\rho(\alpha)}$. Then σ is plainly monotone and, as we show, cofinal. For, given α, β with $\beta \geqslant \rho(\alpha)$, we use

the cofinality of ρ to infer $\alpha' \geqslant \alpha$ with $\rho(\alpha') \geqslant \beta$; then $(\alpha', \rho(\alpha')) \geqslant (\alpha, \beta)$. It is now obvious that Θ is a morphism of $\mathfrak{C}^S$.

We define $\bar{\Theta} = (\bar{\sigma}, 1.)$, where $\bar{\sigma}(\alpha, \beta) = \alpha$ and $1_{\alpha\beta} = 1 : \bar{C}_{\alpha\beta} \to C_\alpha$. Again it is clear that $\bar{\Theta}$ is a morphism of $\mathfrak{C}^S$ and, moreover,

$$\bar{\Theta}\Theta = 1 : (A;C.) \to (A;C.) . \tag{2.4}$$

Finally we define $\bar{\Phi} = (\bar{\rho}, \bar{\varphi}.)$, where $\bar{\rho}(\alpha, \beta) = \beta$ and $\bar{\varphi}_{\alpha\beta} = g_{\rho(\alpha),\beta}\varphi_\alpha$. It is again clear that $\bar{\rho}$ is monotone and cofinal, but, in fact, $\bar{\rho}$ *is a fibre-map*. For, given α, β with $\beta \geqslant \rho(\alpha)$ and given $\beta' \geqslant \beta$, then $(\alpha, \beta') \in \bar{A}$ and $\bar{\rho}(\alpha, \beta') = \beta'$, $(\alpha, \beta') \geqslant (\alpha, \beta)$. Moreover, $\bar{\varphi}.$ satisfies the required commutativity relation. For if $(\alpha, \beta), (\alpha', \beta') \in \bar{A}$ with $(\alpha, \beta) \leqslant (\alpha', \beta')$ then we have the diagram

$$\begin{tikzcd}
C_\alpha \arrow[r, "f_{\alpha\alpha'}"] \arrow[d, "\varphi_\alpha"'] & C_{\alpha'} \arrow[d, "\varphi_{\alpha'}"] \\
D_{\rho(\alpha)} \arrow[r, "g_{\rho(\alpha),\rho(\alpha')}"'] \arrow[d, "g_{\rho(\alpha),\beta}"'] & D_{\rho(\alpha')} \arrow[d, "g_{\rho(\alpha'),\beta'}"] \\
D_\beta \arrow[r, "g_{\beta\beta'}"'] & D_{\beta'}
\end{tikzcd}$$

where the top square commutes by (2.1) and the bottom square commutes because $D.$ is a direct system over B.

This completes the description of the standard factorization, for, plainly,

$$\bar{\Phi}\Theta = \Phi . \tag{2.5}$$

We next relate the notions of *fibre-map* and *pull-back*. First we place ourselves in the category $\mathfrak{D}\mathfrak{S}$ of directed sets and monotone functions. Let

$$\begin{tikzcd}
 & A \arrow[d, "\rho"] \\
B \arrow[r, "\sigma"'] & X
\end{tikzcd} \tag{2.6}$$

be a diagram in $\mathfrak{D}\mathfrak{S}$ and let

$$\begin{tikzcd}
Y \arrow[r, "\rho'"] \arrow[d, "\sigma'"'] & A \arrow[d, "\rho"] \\
B \arrow[r, "\sigma"'] & X
\end{tikzcd} \tag{2.7}$$

be the pull-back of (2.6) in the underlying category of sets. Then Y may be given

the structure of an ordered set in the obvious way and ρ', σ' are then monotone functions.

Theorem 2.8. *Suppose that σ in (2.6) is a fibre-map. Then*

(a) *Y is a directed set;*

(b) *ρ' is a fibre-map;*

(c) *σ' is a morphism of $\mathfrak{D}\mathfrak{S}$;*

(d) *(2.7) is a pull-back in $\mathfrak{D}\mathfrak{S}$.*

Proof: (a) Let $(\alpha_i, \beta_i) \in Y$, $i = 1, 2$; thus $\rho(\alpha_i) = \sigma(\beta_i)$. Choose $\beta' \geqslant \beta_1, \beta_2$ and then choose $\alpha \geqslant \alpha_1, \alpha_2$ such that $\rho(\alpha) \geqslant \sigma(\beta')$. Since σ is a fibre-map, we may find $\beta \geqslant \beta'$ with $\sigma(\beta) = \rho(\alpha)$. Then $(\alpha, \beta) \in Y$ and $(\alpha, \beta) \geqslant (\alpha_i, \beta_i)$, $i = 1, 2$.

(b) Let $(\alpha, \beta) \in Y$ and $\alpha' \geqslant \alpha$. Then $\rho(\alpha') \geqslant \rho(\alpha) = \sigma(\beta)$. Since σ is a fibre-map, we may find $\beta' \geqslant \beta$ with $\sigma(\beta') = \alpha'$.

(c) We must show σ' cofinal. Let $\beta \in B$ and choose $\alpha \in A$ with $\rho(\alpha) \geqslant \sigma(\beta)$. Since σ is a fibre-map, there exists $\beta' \geqslant \beta$ with $\sigma(\beta') = \rho(\alpha)$. Then $(\alpha, \beta') \in Y$ and $\sigma'(\alpha, \beta') = \beta' \geqslant \beta$.

(d) Let

$$\begin{tikzcd} \bar{Y} \arrow[r, "\bar{\rho}"] \arrow[d, "\bar{\sigma}"'] & A \arrow[d, "\rho"] \\ B \arrow[r, "\sigma"'] & X \end{tikzcd}$$

be a commutative diagram in $\mathfrak{D}\mathfrak{S}$. Then, since (2.7) is the pull-back in the category of sets, there is a unique function $\tau : \bar{Y} \to Y$ such that $\rho'\tau = \bar{\rho}$, $\sigma'\tau = \bar{\sigma}$. It remains to show that τ is monotone cofinal.

Now τ is defined by $\tau(\eta) = (\bar{\rho}(\eta), \bar{\sigma}(\eta))$, $\eta \in \bar{Y}$. It is thus plain that, $\bar{\rho}$ and $\bar{\sigma}$ being monotone, τ is monotone. Suppose now that $(\alpha, \beta) \in Y$. Then there exist η', η'' in $\bar{Y}$ with $\bar{\sigma}(\eta') \geqslant \alpha$, $\bar{\sigma}(\eta'') \geqslant \beta$. Choose $\eta \geqslant \eta', \eta''$. Then $\bar{\rho}(\eta) \geqslant \alpha$, $\bar{\sigma}(\eta) \geqslant \beta$, so $\tau(\eta) \geqslant (\alpha, \beta)$ and τ is cofinal. This completes the proof of the theorem.

Next suppose given a diagram in $\mathfrak{C}^{\mathrm{S}}$

$$\begin{tikzcd} & (A; C.) \arrow[d, "\Phi = (\rho, \varphi.)"] \\ (B; D.) \arrow[r, "\Psi = (\sigma, \psi.)"'] & (X; E.) \end{tikzcd}$$

where Ψ is a fibre-map and suppose that $\mathfrak{C}$ admits pull-backs. Then, for each $\gamma = (\alpha, \beta) \in Y$ we may construct the pull-back, in $\mathfrak{C}$,

(2.10)
$$\begin{tikzcd} F_\gamma \arrow[r, "\varphi'_\gamma"] \arrow[d, "\psi'_\gamma"'] & C_\gamma \arrow[d, "\varphi_\alpha"] \\ D_\beta \arrow[r, "\psi_\beta"'] & E_{\rho(\alpha)} \end{tikzcd}$$

Moreover if $\gamma' \geqslant \gamma$, $\gamma' = (\alpha', \beta')$, a straightforward argument yields a morphism $k_{\gamma\gamma'} : F_\gamma \to F_{\gamma'}$ such that

(2.11) $\qquad \varphi'_{\gamma'} k_{\gamma\gamma'} = f_{\alpha\alpha'} \varphi'_\gamma, \qquad \psi'_{\gamma'} k_{\gamma\gamma'} = g_{\beta\beta'} \psi'_\gamma$;

and $F. = \{F_\gamma; k_{\gamma\gamma'}\}$ is then a direct system in $\mathfrak{C}$ over Y, and $\Phi' = (\rho', \varphi'.)$, $\Psi' = (\sigma', \psi'.)$ are then morphisms $(Y;F.) \to (A;C.)$, $(Y;F.) \to (B;D.)$ in $\mathfrak{C}^S$ with Φ' a fibre-map.

Theorem 2.12. *The diagram*

$$\begin{tikzcd} (Y;F.) \arrow[r, "\Phi'"] \arrow[d, "\Psi'"'] & (A;C.) \arrow[d, "\Phi"] \\ (B;D.) \arrow[r, "\Psi"'] & (X;E.) \end{tikzcd}$$

is a pull-back in $\mathfrak{C}^S$.

Proof: The diagram obviously commutes. Suppose now that we have the commutative diagram

$$\begin{tikzcd} (\bar{Y};\bar{F}.) \arrow[r, "\bar{\Phi}"] \arrow[d, "\bar{\Psi}"'] & (A;C.) \arrow[d, "\Phi"] \\ (B;D.) \arrow[r, "\Psi"'] & (X;E.) \end{tikzcd}, \quad \bar{\Phi} = (\bar{\rho}, \bar{\varphi}.), \quad \bar{\Psi} = (\bar{\sigma}, \bar{\psi}.).$$

Then, as shown in the proof of Theorem 2.8, we have a unique $\tau : \bar{Y} \to Y$ such that $\rho'\tau = \bar{\rho}$, $\sigma'\tau = \bar{\sigma}$. Given $\eta \in \bar{Y}$, consider the diagram

$$\begin{tikzcd} \bar{F}_\eta \arrow[drr, "\bar{\varphi}_\eta"] \arrow[ddr, "\bar{\psi}_\eta"'] & & \\ & F_{\tau(\eta)} \arrow[r, "\varphi'_{\tau(\eta)}"'] \arrow[d, "\psi'_{\tau(\eta)}"] & C_{\bar{\rho}(\eta)} \arrow[d, "\varphi_{\bar{\rho}(\eta)}"] \\ & D_{\bar{\sigma}(\eta)} \arrow[r, "\psi_{\bar{\sigma}(\eta)}"'] & E_{\rho\bar{\rho}(\eta)} \end{tikzcd}$$

Since (2.10) is a pull-back in $\mathfrak{C}$, there is a unique $\theta_\eta : F_\eta \to F_{\tau(\eta)}$ such that

(2.14) $$\varphi'_{\tau(\eta)}\theta_\eta = \bar{\varphi}_\eta, \quad \psi'_{\tau(\eta)}\theta_\eta = \bar{\psi}_\eta \ .$$

Again a straightforward argument establishes the commutativity of the diagram

$$\begin{tikzcd} \bar{F}_\eta \arrow[r, "\bar{k}_{\eta\eta'}"] \arrow[d, "\theta_\eta"] & \bar{F}_{\eta'} \arrow[d, "\theta_{\eta'}"] \\ F_{\tau(\eta)} \arrow[r, "k_{\tau(\eta),\tau(\eta')}"'] & F_{\tau(\eta')} \end{tikzcd}$$

where $\eta \leqslant \eta'$ and $\bar{F}. = \{\bar{F}_\eta ; \bar{k}_{\eta\eta'}\}$. Thus $T = (\tau ; \theta.) : (\bar{Y}; \bar{F}.) \to (Y; F.)$ is the unique morphism such that

$$\Phi' T = \bar{\Phi}, \quad \Psi' T = \bar{\Psi}$$

and the theorem is proved.

We now turn our attention to the direct limit functor. We assume $\mathfrak{C}$ cocomplete so that, for any directed set A, we may form the category $\mathfrak{C}^A$ of direct systems in $\mathfrak{C}$ over A and we have the direct limit

$$\varinjlim^A : \mathfrak{C}^A \to \mathfrak{C} \ .$$

Now let $\rho : A \to B$ be a morphism of $\mathfrak{T}\mathfrak{S}$; then ρ induces the functor

$$\rho^* : \mathfrak{C}^B \to \mathfrak{C}^A$$

given by $\rho^* \{C_\beta ; g_{\beta\beta'}\} = \{C_\alpha ; g_{\alpha\alpha'}\}$, where

$$C_\alpha = C_{\rho(\alpha)}, \quad g_{\alpha\alpha'} = g_{\rho(\alpha)\rho(\alpha')} \ .$$

The following theorem appears in many places with varying degrees of generality (see e.g. * [2], Theorem 3.7); we suppress the proof.

Theorem 2.15. $\varinjlim^B = \varinjlim^A \circ \rho^*$.

* In this reference, we allow more general index categories but ρ is one-one; this latter restriction is, however, quite unnecessary to the argument.

We now define a functor $\varinjlim^{S} : \mathfrak{C}^{S} \to \mathfrak{C}$ by the rule

$$\varinjlim^{S}(A ; C_\alpha) = \varinjlim^{A}(C_\alpha) \quad ,$$

and for $\Phi = (\rho, \varphi.) : (A ; C_\alpha) \to (B ; D_\beta)$,

$$\varinjlim^{S}\Phi = \varinjlim^{A}(\varphi^{A}_{\alpha}) \, ,$$

where (φ^{A}_{α}) is the evident morphism $(C_\alpha) \to (\rho^*(D_\beta))$ in $\mathfrak{C}^{A}$, that is, $\varphi^{A}_{\alpha} : C_\alpha \to D_\alpha$ is just $\varphi_\alpha : C_\alpha \to D_{\rho(\alpha)}$.

Theorem 2.16. $\varinjlim^{S}$ *is a functor left-adjoint, left-inverse to the embedding functor* $\vec{P} : \mathfrak{C} \to \mathfrak{C}^{S}$.

Proof: That $\varinjlim^{S}$ is a functor follows from Theorem 2.15 together with the evident relation

$$(2.17) \qquad \theta^{A}_{\alpha} = (\rho^* \psi^{B}_{\beta}) \circ \varphi^{A}_{\alpha} \, ,$$

where $(\sigma, \psi.)(\rho, \varphi.) = (\sigma\rho, \theta.)$.

Let us write L^{A}, L^{S} for $\varinjlim^{A}$, $\varinjlim^{S}$ and P^{A}, P^{S} for the constant embeddings $P^{A} : \mathfrak{C} \to \mathfrak{C}^{A}$, $P^{S} : \mathfrak{C} \to \mathfrak{C}^{S}$ (recall that $P^{S} = \vec{P}$ was defined at the beginning of this section). Then. $L^{A}P^{A} = 1$ and there is a natural transformation $\pi^{A} : 1 \to P^{A}L^{A}$ and

$$(2.18) \qquad L^{A}\pi^{A} = 1, \quad \pi^{A}P^{A} = 1 \, , \quad \text{all } A.$$

It is obvious that $L^{S}P^{S} = 1$. We define $\pi^{S} : 1 \to P^{S}L^{S}$ by

$$\pi^{S}(A ; C_\alpha) = \Pi^{A} \; : (A ; C_\alpha) \to (O ; L^{A}(C_\alpha)) \, ,$$

where O is the singleton set and $\Pi^{A} = (O, \pi^{A}_{\cdot})$, so that $\pi^{A}_{\alpha} : C_\alpha \to L^{A}(C_\alpha)$ is the morphism given by π^{A}. It is then obvious that (2.18) imply

$$(2.19) \qquad L^{S}\pi^{S} = 1, \quad \pi^{S}P^{S} = 1 \; ,$$

and the theorem is proved. We may express this theorem simply by saying that the direct limits in each $\mathfrak{C}^{A}$ yield a direct limit in $\mathfrak{C}^{S}$.

We close this section with an elementary remark about the relation of direct limits to the standard factorization (2.3). Let $\Phi : (A ; C.) \to (B ; D.)$ be a morphism in $\mathfrak{C}^{S}$ with $\Phi = (\rho, \varphi.)$ where $\varphi_\alpha : C_\alpha \to D_{\rho(\alpha)}$ is invertible for each $\alpha \in A$ so that $(\varphi^{A}_{\alpha}) : (C_\alpha) \to \rho^*(D_\beta)$ is invertible in $\mathfrak{C}^{A}$. It follows that $\varinjlim^{S}\Phi$ *is invertible in* $\mathfrak{C}$. In particular, we have in (2.3)

$$(2.20) \qquad \varinjlim^{S}\Theta = 1 \, , \qquad \varinjlim^{S}\bar{\Theta} = 1, \quad \varinjlim^{S}\Phi = \varinjlim^{S}\bar{\Phi} \; .$$

§ 3. Direct systems of groups

Let $\mathfrak{G}$ be the category of groups and let $\mathfrak{G}_0 \subseteq \mathfrak{G}_1$ be full subcategories of $\mathfrak{G}$ such that the objects of $\mathfrak{G}_1$ are precisely the images of objects of $\mathfrak{G}_0^{\mathrm{S}}$ under the direct limit functor

$$\varinjlim : \mathfrak{G}_0^{\mathrm{S}} \to \mathfrak{G} .$$

We make the following hypothesis throughout this section.

Hypothesis 3.1. If

$$\begin{array}{lc} (3.2) & \begin{array}{ccc} & & G_0 \\ & & \downarrow \\ G_0' & \longrightarrow & G_1 \end{array} \end{array}$$

is a diagram in $\mathfrak{G}$, with $G_0, G_0' \in |\mathfrak{G}_0|$, $G_1 \in |\mathfrak{G}_1|$, then the ($\mathfrak{G}$-) pull-back of (3.2) is in $\mathfrak{G}_0$.

Now we may pass to the *category of fractions* relative to the direct limit functor. We denote this by $\vec{\mathfrak{G}}_0^{\mathrm{S}}$. There is an embedding functor $i : \mathfrak{G}_0^{\mathrm{S}} \to \vec{\mathfrak{G}}_0^{\mathrm{S}}$ and $\varinjlim$ extends to a functor

$$(3.3) \qquad L : \vec{\mathfrak{G}}_0^{\mathrm{S}} \to \mathfrak{G}$$

such that $Li = \varinjlim$. Recall that we obtain $\vec{\mathfrak{G}}_0^{\mathrm{S}}$ from $\mathfrak{G}_0^{\mathrm{S}}$ by adjoining to the latter formal inverses of those morphisms Φ such that $\varinjlim \Phi$ is an isomorphism. Thus a morphism of $\vec{\mathfrak{G}}_0^{\mathrm{S}}$ from $(A;C.)$ to $(B;D.)$ is represented by a *path*

$$(3.4) \qquad (A^1;C^1_.) \leftarrow (A^2;C^2_.) \to \dots \leftarrow (A^{n-1};C^{n-1}_.) \to (A^n;C^n_.)$$

where $(A^1;C^1_.) = (A;C.)$, $(A^n;C^n_.) = (B;D.)$, the arrows from left to right are morphisms of $\mathfrak{G}_0^{\mathrm{S}}$ and the arrows from right to left are morphisms Φ of $\mathfrak{G}_0^{\mathrm{S}}$ such that $\varinjlim \Phi$ is an isomorphism.

Our main theorem is the following: recall that we assume Hypothesis 3.1.

Theorem 3.5. *The functor* (3.3)

$$L : \vec{\mathfrak{G}}_0^{\mathrm{S}} \to \mathfrak{G}$$

is full and faithful.

The proof depends on the following lemma, certainly well-known if we replace $\mathfrak{G}^{\mathrm{S}}$ by $\mathfrak{G}^{A}$.

Lemma 3.6. *The direct limit in $\mathfrak{G}^S$ commutes with pull-backs of fibre-maps.*

The lemma means the following. Suppose given, in $\mathfrak{G}^S$, the diagram

(3.7)
$$\begin{tikzcd}
(A;C.) \arrow[rrr, "L"] \arrow[dr, "\Phi = (\rho, \varphi.)"] & & & C \arrow[dr, "\varphi"] & \\
& (X;E.) \arrow[r, "L"] & F \arrow[ur, "\varphi'"] \arrow[dr, "\psi'"'] \arrow[rr] & & E \\
(B;D.) \arrow[ur, "\Psi = (\sigma, \psi.)"'] \arrow[rrr, "L"'] & & & D \arrow[ur, "\psi"'] &
\end{tikzcd}$$

where (i) the square on the right is a pull-back, (ii) the horizontal arrows are the direct limit functor with

$$L\Phi = \varphi, \quad L\Psi = \psi.$$

(iii) Ψ is a fibre-map.

We may then construct the pull-back of Φ and Ψ as in §2,

(3.8)
$$\begin{tikzcd}
& (A;C.) \arrow[dr, "\Phi"] & \\
(Y;F.) \arrow[ur, "\Phi'"] \arrow[dr, "\Psi'"'] & & (X;E.) \\
& (B;D.) \arrow[ur, "\Psi"'] &
\end{tikzcd}$$

and the lemma claims that $L(Y;F.) = F$, $L\Phi' = \varphi'$, $L\Psi' = \psi'$.

Proof of lemma: Recall that the direct limit is constructed from the system (C_α) of groups over A by setting up the obvious equivalence relation on $\bigcup_{\alpha \in A} C_\alpha$ and defining in the resulting set C of equivalence classes the obvious group structure. Our task is simply to define, for each $\gamma \in Y$, a homomorphism $\omega_\gamma : F_\gamma \to F$ such that

(a) $\omega_\gamma = \omega_{\gamma'} k_{\gamma\gamma'}$, if $\gamma \leqslant \gamma'$;

(b) every $z \in F$ is in the image of some ω_γ;

(c) if $\omega_\gamma(a) = e$, then $k_{\gamma\gamma'}(a) = e$ for some $\gamma' \geqslant \gamma$, $a \in F_\gamma$.

We define ω_γ in the obvious way: an element of F_γ, $\gamma = (\alpha, \beta)$ is a pair (x, y), $x \in C_\alpha$, $y \in D_\beta$ with $\rho(\alpha) = \sigma(\beta)$, $\varphi_\alpha(x) = \psi_\beta(y)$ (see (2.10)). If $[x]$, $[y]$ are the equivalence classes in C, D of x, y, respectively, then $\varphi[x] = [\varphi_\alpha(x)] = [\psi_\beta(y)] = \psi[y]$, so $([x], [y]) \in F$ and we set

$$\omega_\gamma(x, y) = ([x], [y]).$$

Then ω_γ is plainly a homomorphism. Also $k_{\gamma\gamma'}(x, y)$ is just $(f_{\alpha\alpha'}(x), g_{\beta\beta'}(y))$ so (a) is plainly satisfied. We now prove (b). Let $z = ([x], [y])$, $x \in C_\alpha$, $y \in D_\beta$, $\varphi[x] = \psi[y]$. Choose $\alpha' \geqslant \alpha$ with $\rho(\alpha') \geqslant \sigma(\beta)$ and then, using the fact that σ is a fibre-map, choose $\beta' \geqslant \beta$ with $\sigma(\beta') = \rho(\alpha')$. Since we may replace x, y by $x' = f_{\alpha\alpha'}(x)$, $y' = g_{\beta\beta'}(y)$, there is no real loss of generality in supposing that, originally, $\rho(\alpha) = \sigma(\beta)$. Then $[\varphi_\alpha(x)] = [\psi_\beta(y)]$ and $\varphi_\alpha(x), \psi_\beta(y) \in E_{\rho(\alpha)}$. Thus there exists $\xi \geqslant \rho(\alpha)$ such that

$$h_{\rho(\alpha),\xi}\varphi_\alpha(x) = h_{\rho(\alpha),\xi}\psi_\beta(y),$$

where $E. = \{E_\xi; h_{\xi\xi'}\}$. Choose $\beta'' \geqslant \beta$ so that $\sigma(\beta'') \geqslant \xi$ and choose $\alpha' \geqslant \alpha$ so that $\rho(\alpha') \geqslant \sigma(\beta'')$. Now choose $\beta' \geqslant \beta''$ so that $\sigma(\beta') = \rho(\alpha')$. Then

$$h_{\rho(\alpha),\rho(\alpha')}\varphi_\alpha(x) = h_{\sigma(\beta),\sigma(\beta')}\psi_\beta(y).$$

Thus if $x' = f_{\alpha\alpha'}(x)$, $y' = g_{\beta\beta'}(y)$, then $[x'] = [x]$, $[y'] = [y]$, $\varphi_{\alpha'}(x') = \psi_{\beta'}(y')$, so $(x', y') \in F_{\gamma'}$, $\gamma' = (\alpha', \beta')$. This proves (b).

To prove (c), let $a = (x, y)$, $x \in C_\alpha$, $y \in D_\beta$, $\rho(\alpha) = \sigma(\beta)$, $\varphi_\alpha(x) = \psi_\beta(y)$. Since $[x] = e$, $[y] = e$, there exist $\alpha' \geqslant \alpha$, $\beta' \geqslant \beta$ such that $f_{\alpha\alpha'}(x) = e$, $g_{\beta\beta'}(y) = e$. Now choose $\alpha'' \geqslant \alpha'$ such that $\rho(\alpha'') \geqslant \sigma(\beta')$ and then choose $\beta'' \geqslant \beta'$ such that $\sigma(\beta'') = \rho(\alpha'')$. Then $\gamma'' = (\alpha'', \beta'') \in Y$ and $k_{\gamma\gamma''}(a) = e$.

Plainly $L\Phi' = \varphi'$, $L\Psi' = \psi'$ and so the lemma is completely proved.

Proof of Theorem 3.5. We first show that L is full. We suppose $(A; C.)$, $(B; D.)$ in $\mathfrak{G}_0^S$ with $L(A; C.) = C$, $L(B; D.) = D$ and $\varphi : C \to D$ in $\mathfrak{G}_1$. We consider the diagram

$$\begin{array}{ccc} & & A \\ & & \downarrow \\ B & \longrightarrow & O \end{array}$$

where O is the singleton set, and pull-back to obtain $Y = A \times B$, since obviously a function to a singleton is a fibre-map. We then consider the diagram

$$(3.9) \qquad \begin{array}{ccc} & & (A; C.) \\ & & \downarrow{\scriptstyle (0;\lambda.)} \\ (B; D.) & \xrightarrow{(0;\mu.)} & (O; D) \end{array}$$

where $\lambda_\alpha(x) = \varphi[x]$, $\mu_\beta(y) = [y]$, $x \in C_\alpha$, $y \in D_\beta$. We then pull back (3.9) to obtain

$$(3.10) \qquad \begin{array}{ccc} (Y; F.) & \xrightarrow{\Psi} & (A; C.) \\ \downarrow{\scriptstyle \Phi} & & \downarrow{\scriptstyle (0;\lambda.)} \\ (B; D.) & \xrightarrow{(0;\mu.)} & (O; D) \end{array}$$

By Hypothesis 3.1, $(Y;F.)$ is an object of $\mathfrak{G}_0^S$. By Lemma 3.6, we obtain from (3.10), on passing to the direct limit,

$$\begin{tikzcd} C \arrow[r, "1"] \arrow[d, "\varphi"] & C \arrow[d, "\varphi"] \\ D \arrow[r, "1"] & D \end{tikzcd} \tag{3.11}$$

since (3.11) *is* a pull-back. Thus Ψ is invertible in $\vec{\mathfrak{G}}_0^S$ and $L(\Phi\Psi^{-1}) = \varphi$. We call the path $\Phi\Psi^{-1}$ constructed above the ***minimal*** ***φ-path*** from $(A;C.)$ to $(B;D.)$.

Next we show that L is faithful. Our first step is to show how every path (3.4) may be shortened. Thus let us suppose given

$$\begin{tikzcd} (A;C.) \arrow[r, "L"] \arrow[d, "\Phi"] & C \arrow[d, "\varphi"] \\ (X;E.) \arrow[r, "L"] & E \\ (B;D.) \arrow[r, "L"] \arrow[u, "\Psi"] & D \arrow[u, "\psi"] \end{tikzcd} \tag{3.12}$$

where ψ is an isomorphism and $\psi\theta = \varphi$, $\theta : C \to D$. Then $L(\Psi^{-1}\Phi) = \theta$. If Ψ were a fibre-map we could apply Lemma 3.6 (and Hypothesis 3.1) to infer a pull-back square in $\mathfrak{G}_0^S$

$$\begin{tikzcd} (Y;F.) \arrow[r, "\Phi'"] \arrow[d, "\Psi'"] & (A;C.) \arrow[d, "\Phi"] \\ (B;D.) \arrow[r, "\Psi"'] & (X;E.) \end{tikzcd}$$

whose direct limit is the pull-back square

$$\begin{tikzcd} F \arrow[r, "\varphi'"] \arrow[d, "\psi'"] & C \arrow[d, "\varphi"] \arrow[dl, dashed, "\theta"] \\ D \arrow[r, "\psi"'] & E \end{tikzcd}$$

with φ', therefore, an isomorphism, so that $\theta\varphi' = \psi'$ and Φ' is invertible in $\vec{\mathfrak{G}}_0^S$; moreover

$$\Psi^{-1}\Phi = \Psi'\Phi'^{-1}$$

in $\vec{\mathfrak{G}}_0^S$. This process, applied successively, would clearly enable us to shorten (3.4)

to a path of length $n = 2$; however, the validity of the process appears to depend on Ψ being a fibre-map.

We now show that, in fact, this is not so. We validate the process by invoking the standard factorization (2.3). Given (3.12) without the additional property that Ψ is a fibre-map, we apply the standard factorization to Ψ. Thus we obtain

$$\text{(3.13)}\qquad \begin{tikzcd} & (A;C.) \arrow[d, "\Phi"] \\ (\bar{B};\bar{D}.) \arrow[r, "\bar{\Psi}"] \arrow[d, shift left, "\bar{\Theta}"] & (X;E.) \\ (B;D.) \arrow[u, shift left, "\Theta"] \arrow[ur, "\Psi"'] & \end{tikzcd}, \quad \bar{\Psi} \text{ fibre-map}, \ \bar{\Psi}\Theta = \Psi, \ \bar{\Theta}\Theta = 1 .$$

It immediately follows that $L(\Theta) = 1$, $L(\bar{\Theta}) = 1$, $L(\bar{\Psi}) = L(\Psi)$, so that Θ, $\bar{\Theta}$ and $\bar{\Psi}$ are invertible in $\mathfrak{G}_0^S$ and $\Theta^{-1} = \bar{\Theta}$. We apply the pull-back argument already given above (with $\bar{\Psi}$ now replacing Ψ) to obtain

$$\text{(3.14)}\qquad \begin{tikzcd} (Y;F.) \arrow[r, "\Phi'"] \arrow[d, "\bar{\Psi}'"] & (A;C.) \arrow[d, "\Phi"] \\ (\bar{B};\bar{D}.) \arrow[r, "\bar{\Psi}"] \arrow[d, shift left, "\bar{\Theta}"] & (X;E.) \\ (B;D.) \arrow[u, shift left, "\Theta"] \arrow[ur, "\Psi"'] & \end{tikzcd}$$

Then, setting $\Psi' = \bar{\Theta}\bar{\Psi}'$, we have

$$\Psi^{-1}\Phi = \Theta^{-1}\bar{\Psi}^{-1}\Phi = \bar{\Theta}\bar{\Psi}'\Phi'^{-1} = \Psi'\Phi'^{-1} ,$$

so that the shortening process works quite generally.

We have now proved that any morphism of $\mathfrak{G}_0^S$ from $(A;C.)$ to $(B;D.)$, say, inducing $\varphi : C \to D$ in the limit, may be written in the form

$$\text{(3.15)}\qquad (A;C.) \xleftarrow{\ \bar{\Psi}\ } (\bar{Y};\bar{F}.) \xrightarrow{\ \bar{\Phi}\ } (B;D.)$$

where $L(\bar{\Psi}) = \bar{\psi}$ is an isomorphism, $L(\bar{\Phi}) = \bar{\varphi}$ and $\bar{\varphi}\bar{\psi}^{-1} = \varphi$. We will prove that (3.15) represents the same morphism $\mathfrak{G}_0^S$ as the *minimal φ-path* constructed in proving that L is full; this will establish the fact that L is faithful.

Now (compare (3.9), (3.10)) the diagram

$$\begin{tikzcd} (\bar{Y};\bar{F}.) \arrow[r, "\bar{\Psi}"] \arrow[d, "\bar{\Phi}"] & (A;C.) \arrow[d, "(0;\lambda.)"] \\ (B;D.) \arrow[r, "(0;\mu.)"'] & (O;D) \end{tikzcd}$$

commutes; for if $\gamma \in \bar{Y}$ and $x \in \bar{F}_\gamma$, then $\varphi[\bar{\psi}_\gamma(x)] = \varphi\bar{\psi}[x] = \bar{\varphi}[x] = [\bar{\varphi}_\gamma(x)]$. Thus there exists a unique morphism $T : (\bar{Y}; \bar{F}.) \to (Y; F.)$ such that

$$(3.16) \qquad \Phi T = \bar{\Phi}, \quad \Psi T = \bar{\Psi}.$$

From the second of these equations we get, by passing to the limit, that $L(T) = \bar{\psi}$ so that T is invertible in $\vec{\mathfrak{G}}_0^S$. Thus (3.16) yields

$$\bar{\Phi}\bar{\Psi}^{-1} = \Phi T T^{-1} \Psi^{-1} = \Phi \Psi^{-1},$$

and the theorem is completely proved.

Remark. In a *calculus of fractions* (see [3]) one assumes as an axiom that, given

$$\begin{tikzcd} & {} \arrow[d] \\ {} \arrow[r] & {} \end{tikzcd}$$

with the horizontal arrow 'invertible', one may find a commutative square

$$\begin{tikzcd} {} \arrow[r, dashed] \arrow[d, dashed] & {} \arrow[d] \\ {} \arrow[r] & {} \end{tikzcd}$$

with the dotted horizontal arrow also 'invertible'. We do not have this situation in $\mathfrak{G}_0^S$. For the axiom requires that if we form a category of fractions by formally inverting certain morphisms of $\mathfrak{C}$, then the square should commute *in* $\mathfrak{C}$. In our situation we obtained from

$$\begin{tikzcd} & (A; C.) \arrow[d, "\Phi"] \\ (B; D.) \arrow[r, "\Psi"] & (X; E.) \end{tikzcd}$$

a square (see (3.14))

$$\begin{tikzcd} (Y; F.) \arrow[r, "\Phi'"] \arrow[d, "\Psi'"] & (A; C.) \arrow[d, "\Phi"] \\ (B; D.) \arrow[r, "\Psi"'] & (X; E.) \end{tikzcd}$$

with Φ' invertible, which *commutes in* $\vec{\mathfrak{G}}_0^S$ *but not in* $\mathfrak{G}_0^S$. On the other hand, our category of fractions has a simplifying feature not present in general, namely, the existence of minimal paths.

We give one consequence of Theorem 3.5. We first observe the following elementary fact.

Proposition 3.17. *Let $L : \mathfrak{A} \to \mathfrak{B}$ be a functor which is full and faithful and surjective on objects. Then given a functor $F_0 : \mathfrak{A} \to \mathfrak{C}$, there is a unique functor $F : \mathfrak{B} \to \mathfrak{C}$ with $FL = F_0$.*

Proof. Suppose $L(A_1) = L(A_2) = B$. Then, L being full and faithful, there are unique morphisms $\alpha : A_1 \to A_2$, $\alpha' : A_2 \to A_1$ such that $L(\alpha) = L(\alpha') = 1_B$ and $\alpha'\alpha = 1_{A_1}$, $\alpha\alpha' = 1_{A_2}$. Thus any two pre-images of B are *canonically* equivalent, so that there is no ambiguity in setting $F = F_0 L^{-1}$ on objects and morphisms. F is plainly uniquely determined by the equation $FL = F_0$ and is evidently a functor.

Theorem 3.18. *Let $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \subseteq \mathfrak{G}$ be as at the outset of this section. Let $\mathfrak{C}$ be a cocomplete category and let $F_0 : \mathfrak{G}_0 \to \mathfrak{C}$ be a functor with the property that, for any morphism Φ in $\mathfrak{G}_0^S$, $\varinjlim F_0^S(\Phi)$ is an equivalence if $\varinjlim \Phi$ is an equivalence (isomorphism). Then, granted Hypothesis 3.1, F_0 extends to a unique functor $F_1 : \mathfrak{G}_1 \to \mathfrak{C}$ such that*

$$(3.19)\qquad F_1 \varinjlim = \varinjlim F_0^S .$$

Proof. Consider the diagram

$$(3.20)\qquad \begin{tikzcd} \mathfrak{G}_0^S \arrow[r, "F_0^S"] \arrow[d, "i"] & \mathfrak{C}^S \arrow[d, "i"] \\ \mathfrak{G}_0^{\vec{S}} \arrow[r, dashed, "\vec{F}_0^S"] \arrow[d, "L"] & \mathfrak{C}^{\vec{S}} \arrow[d, "L"] \\ \mathfrak{G}_1 \arrow[r, dashed, "F_1"] & \mathfrak{C} \end{tikzcd}$$

The property demanded of F_0 is precisely that which yields a (unique) functor $\vec{F}_0^S$ of the category of fractions $\mathfrak{G}_0^{\vec{S}}$ into the category of fractions $\mathfrak{C}^{\vec{S}}$. We now apply Proposition 3.17 (with F_0 in that proposition replaced by $L\vec{F}_0^S$) to obtain F_1. It is plain that (3.19) does determine F_1 uniquely, since, as we showed in proving Theorem 3.5, given a morphism φ in $\mathfrak{G}_1$ there does exist a morphism Φ in $\mathfrak{G}_0^S$ with $\varinjlim \Phi = \varphi$ (namely, the Φ of (3.10)).

We remark that (3.19) obviously implies that $\varinjlim F_0^S(\Phi)$ is an equivalence if $\varinjlim \Phi$ is an equivalence, so that the force of the theorem is that this necessary condition for the existence of F_1 is also a sufficient condition. An application of Theorem 3.18 will be found in [6].

§ 4. The category of finitely-generated groups

We would wish to be able to apply Theorem 3.18 to the case in which $\mathfrak{G}_0$ is the category of finitely-generated groups, so that $\mathfrak{G}_1 = \mathfrak{G}$. However this is not possible as the theorem stands since Hypothesis 3.1 would be violated in this case. On the other hand we do then have the 'dual' hypothesis, namely

Hypothesis 4.1. Let $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \subseteq \mathfrak{G}_2$ with $\mathfrak{G}_2$ complete and cocomplete. If

$$\begin{tikzcd} G_2 \arrow[r] \arrow[d] & G_0 \\ G_0' & \end{tikzcd} \tag{4.2}$$

is a diagram in $\mathfrak{G}$ with $G_0, G_0' \in |\mathfrak{G}_0|$, $G_2 \in |\mathfrak{G}_2|$, then the ($\mathfrak{G}_2$-) push-out of (4.2) is in $\mathfrak{G}_0$.

Thus, in practice, $\mathfrak{G}_2 = \mathfrak{G}$ or Ab, the category of Abelian groups. As remarked, Hypothesis 4.1 holds if $\mathfrak{G}_0$ is the category of finitely-generated groups and $\mathfrak{G}_1 = \mathfrak{G}_2 = \mathfrak{G}$.

Our object is to prove a theorem similar to Theorem 3.5 under Hypothesis 4.1. We make a preliminary observation. Suppose $(A;C.) \in |\mathfrak{G}^S|$ and $\varinjlim (A;C.) = C$. Let B be any directed set. Then $A \times B$ is also a directed set and we may define the object $(A \times B; \bar{C}.)$ of $|\mathfrak{G}^S|$ by

$$\bar{C}_{\alpha\beta} = C_\alpha, \quad \bar{f}_{\alpha\beta,\alpha'\beta'} = f_{\alpha\alpha'} : \bar{C}_{\alpha\beta} \to \bar{C}_{\alpha'\beta'}$$

where, as usual, $C. = \{C_\alpha; f_{\alpha\alpha'}\}$. Moreover, there is an obvious morphism $\Pi : (A \times B; \bar{C}.) \to (A; C.)$ such that

$$\varinjlim \Pi = 1_C . \tag{4.3}$$

We will make heavy use of Π in this section. We may say we obtain $\bar{C}.$ by 'blowing up' the indexing set by means of B.

We next prove a lemma

Lemma 4.4. *Let* $\varinjlim (Y;C.) = C$, $\varinjlim (Y;D.) = D$, *with* $(Y;C.), (Y;D.)$ *in* $\mathfrak{G}_0^S$ *and let* $\varphi : C \to D$. *Then under Hypothesis 4.1, there exists a path*

$$(Y;C.) \xrightarrow{\Gamma} (Y;E.) \xleftarrow{\Delta} (Y;D.)\,, \quad \Gamma = (1,\gamma.)\,, \quad \Delta = (1,\delta.) \tag{4.5}$$

in $\vec{\mathfrak{G}}_0^S$ *such that* $L(\Delta^{-1}\Gamma) = \varphi$, *and all such paths* (4.5) *represent the same morphism of* $\vec{\mathfrak{G}}_0^S$.

Proof. We proceed as in the proof of Theorem 3.5, except that we operate with the fixed directed set Y which may now be omitted from the notation. We replace $(O;D)$ in (3.9) with the constant system (D) over Y so that we obtain the pull-back (3.10)

(4.6)
$$\begin{tikzcd}
(F.) \arrow[r, "\Psi"] \arrow[d, "\Phi"] & (C.) \arrow[d, "(\lambda.)"] \\
(D.) \arrow[r, "(\mu.)"'] & (D)
\end{tikzcd}$$

Now, however, $(F.)$ is in $\mathfrak{G}_2^S$. We now push out Φ and Ψ, using the elementary fact that *push-outs commute with direct limits.* We obtain

(4.7)
$$\begin{tikzcd}
(F.) \arrow[r, "\Psi"] \arrow[d, "\Phi"] & (C.) \arrow[d, "\Gamma."] \arrow[ddr, "(\lambda.)"] & \\
(D.) \arrow[r, "\Delta"] \arrow[drr, "(\mu.)"'] & (E.) \arrow[dr, "(\nu.)"] & \\
 & & (D)
\end{tikzcd}$$

By Hypothesis 4.1, $(E.)$ is in $\mathfrak{G}_0^S$. Moreover, applying L to (4.7) we obtain

$$\begin{tikzcd}
C \arrow[r, "1"] \arrow[d, "\varphi"] & C \arrow[d, "\varphi"] \arrow[ddr, "\varphi"] & \\
D \arrow[r, "1"] \arrow[drr, "1"'] & D \arrow[dr, "1"] & \\
 & & D
\end{tikzcd}$$

so that Δ is invertible and the existence of (4.5) is proved. We call the path we have constructed the *base φ-path* from $(Y;C.)$ to $(Y;D.)$. We now show that any path of the form (4.5) is equivalent in $\mathfrak{G}_0^S$ to the base-path.

Suppose that we are given the φ-path, in $\mathfrak{G}_0^S$,

(4.8)
$$(Y;C.) \xrightarrow{\Gamma'} (Y;E.) \xleftarrow{\Delta'} (Y;D.)\,,$$

$$L(\Delta'^{-1}\Gamma') = \varphi, \quad \Gamma' = (1, \gamma'.)\,, \quad \Delta' = (1, \delta'.)\,.$$

We pull (4.8) back and then push out again, obtaining

(4.9)
$$\begin{tikzcd}
(F'_.) \arrow[r, "\Psi'"] \arrow[d, "\Phi'"] & (C.) \arrow[d, "\Gamma''"] \arrow[ddr, "\Gamma'"] & \\
(D.) \arrow[r, "\Delta''"] \arrow[drr, "\Delta'"'] & (E''_.) \arrow[dr, "\Theta'"] & \\
 & & (E'_.)
\end{tikzcd}$$

with $(E.'')$ again in $\mathfrak{G}_0^S$.
We also note that we have the commutative diagram

(4.10)
$$\begin{tikzcd}
 & (C.) \arrow[d, "\Gamma'"] \arrow[ddr, "(\lambda.)"] & \\
(D.) \arrow[r, "\Delta'"] \arrow[drr, "(\mu.)"'] & (E'_.) \arrow[dr, "(\nu'_.)"] & \\
 & & (D)
\end{tikzcd}$$

where $\nu'_\gamma(x) = \delta'^{-1}[x]$, $\gamma \in Y$, $x \in E'_\gamma$, $\delta' = L(\Delta')$.
Since $\Gamma'\Psi' = \Delta'\Phi'$, it follows that $(\lambda.)\Psi' = (\mu.)\Phi'$, so that there exists $T : (F'.) \to (F.)$ with

(4.11) $\qquad \Phi T = \Phi', \quad \Psi T = \Psi'.$

Thus $\Gamma\Psi' = \Delta\Phi'$ so there exists $\Theta : (E''_.) \to (E.)$ with

(4.12) $\qquad \Theta\Gamma'' = \Gamma, \quad \Theta\Delta'' = \Delta.$

From (4.9) we also have

(4.13) $\qquad \Theta'\Gamma'' = \Gamma', \quad \Theta'\Delta'' = \Delta'.$

By applying $\vec{L}$ to (4.9) we readily conclude that Δ'', and hence also Θ', are invertible in $\mathfrak{G}_0^S$. Then (4.12) shows that Θ is also invertible in $\vec{\mathfrak{G}}_0^S$ and (4.12), (4.13) show that

$$\Delta^{-1}\Gamma = \Delta''^{-1}\Gamma'' = \Delta'^{-1}\Gamma'$$

in $\vec{\mathfrak{G}}_0^S$, proving the lemma.

Suppose now given $(A;C.)$, $(B;D.)$ in $\mathfrak{G}_0^S$ with $\varinjlim (A;C.) = C$, $\varinjlim (B;D.) = D$

and $\varphi : C \to D$. A *standard φ-path* from $(A;C.)$ to $(B;D.)$ is a path of the form

$$(4.14)\qquad (A;C.) \xleftarrow{\ \Pi\ } (A\times B;\bar{C}.) \xrightarrow{\ \Gamma\ } (A\times B;E.) \xleftarrow{\ \Delta\ } (A\times B;\bar{D}.) \xrightarrow{\ \Pi\ } (B;D.),$$

inducing φ in the limit, with Γ and Δ the identity on $A\times B$.

Theorem 4.15. (i) *There exists a unique standard φ-path in $\vec{\mathfrak{G}}^S_0$.*
(ii) *The identity on $(A;C.)$ is the standard 1_c-path.*
(iii) *The composite of the standard φ-path and the standard ψ-path is the standard $\psi\varphi$-path.*

Proof. (i) is contained in Lemma 4.4, together with (4.3). To prove (ii), consider the standard 1-path

$$(A;C.) \xleftarrow{\ \Pi_1\ } (A\times A;C^1.) \xrightarrow{\ \Gamma\ } (A\times A;D.) \xleftarrow{\ \Delta\ } (A\times A;C^2.) \xrightarrow{\ \Pi_2\ } (A;C.),$$

where $C^1_{\alpha\alpha'} = C_\alpha$, $C^2_{\alpha\alpha'} = C_{\alpha'}$. Define $(A;\tilde{D}.)$ by

$$\tilde{D}_\alpha = D_{\alpha\alpha},\quad \tilde{g}_{\alpha\alpha'} = g_{\alpha\alpha,\alpha'\alpha'},\quad \alpha \leqslant \alpha'.$$

There are then *diagonal* morphisms $\nabla_i : (A;C.) \longrightarrow (A\times A;C^i.)$, $i = 1, 2$, and $\nabla : (A;\tilde{D}.) \longrightarrow (A\times A;D.)$ and, by restricting to the diagonal, morphisms

$$\tilde{\Gamma} : (A;C.) \longrightarrow (A;\tilde{D}.),\quad \tilde{\Delta} : (A;C.) \longrightarrow (A;\tilde{D}.)$$

such that the diagram

$$\begin{tikzcd}
(A;C.) & (A\times A;C^1.) \arrow[l, "\Pi_1"'] \arrow[r, "\Gamma"] & (A\times A;D.) & (A\times A;C^2.) \arrow[l, "\Delta"'] \arrow[r, "\Pi_2"] & (A;C.) \\
& (A;C.) \arrow[ul, equal] \arrow[u, "\nabla_1"'] \arrow[r, "\tilde{\Gamma}"] & (A;\tilde{D}.) \arrow[u, "\nabla"'] & (A;C.) \arrow[l, "\tilde{\Delta}"'] \arrow[u, "\nabla_2"'] \arrow[ur, equal] &
\end{tikzcd}$$

commutes; moreover, every morphism in this diagram induces the identity in the limit. Thus the standard 1-path is equal, in $\vec{\mathfrak{G}}^S_0$, to

$$(A;C.) \xrightarrow{\ \tilde{\Gamma}\ } (A;\tilde{D}.) \xleftarrow{\ \tilde{\Delta}\ } (A;C.)$$

and hence, by Lemma 4.4, to the identity on $(A;C.)$. It remains to prove (iii); as a notational convenience we write AB for $A\times B$, etc. Thus we are given the path

$$(4.16)\qquad (A;C.) \xleftarrow{\ \Pi\ } (AB;\bar{C}.) \xrightarrow{\ \Gamma\ } (AB;E.) \xleftarrow{\ \Delta\ } (AB;\bar{D}.) \xrightarrow{\ \Pi\ } (B;D.) \xleftarrow{\ \Pi\ } (BX;\bar{\bar{D}}.)$$
$$\xrightarrow{\ \Gamma'\ } (BX;F.) \xleftarrow{\ \Delta'\ } (BX;\bar{G}.) \xrightarrow{\ \Pi\ } (X;G.)$$

where we may assume $L(\Gamma) = \varphi$, $L(\Delta) = 1$, $L(\Gamma') = \psi$, $L(\Delta') = 1$. Let $Y = ABX$. Then we claim there is an obvious commutative diagram *

(4.17)

$$\begin{tikzcd}
(AB;\bar{C}.) \arrow[r, "\Gamma"] & (AB;E.) & (AB;D.) \arrow[l, "\Delta"'] \arrow[r, "\Pi"] & (B;D.) & (BX;D_1) \arrow[l, "\Pi"'] \arrow[r, "\Gamma'"] & (BX;F.) & (BX;\bar{G}.) \arrow[l, "\Delta'"'] \\
(Y;\tilde{C}.) \arrow[u, "\Pi"] \arrow[r, "\tilde{\Gamma}"] & (Y;\tilde{E}.) \arrow[u, "\Pi"] & (Y;\tilde{D}.) \arrow[u, "\Pi"] \arrow[l, "\tilde{\Delta}"'] \arrow[r, "1"] & (Y;\tilde{D}.) \arrow[u, "\Pi"] & (Y;\tilde{D}.) \arrow[u, "\Pi"] \arrow[l, "1"'] \arrow[r, "\tilde{\Gamma}'"] & (Y;\tilde{F}.) \arrow[u, "\Pi"] & (Y;\tilde{G}.) \arrow[u, "\Pi"] \arrow[l, "\tilde{\Delta}'"']
\end{tikzcd}$$

where $\tilde{C}_{\alpha\beta\gamma} = C_\alpha$, etc., and all the morphisms on the bottom row are the identity on Y. Thus we consider

$$(4.18)\qquad (\tilde{C}.) \xrightarrow{\tilde{\Gamma}} (\tilde{E}.) \xleftarrow{\tilde{\Delta}} (\tilde{D}.) \xrightarrow{\tilde{\Gamma}'} (\tilde{F}.) \xleftarrow{\tilde{\Delta}'} (\tilde{G}.)\,,$$

again suppressing the directed set Y from the notation. We may now take the push-out

(4.19)

$$\begin{tikzcd}
(\tilde{D}.) \arrow[r, "\tilde{\Delta}"] \arrow[d, "\tilde{\Gamma}'"] & (\tilde{E}.) \arrow[d, "\Delta_*"] \\
(\tilde{F}.) \arrow[r, "\Gamma'_*"'] & (\tilde{H}.)
\end{tikzcd}$$

Certainly $(\tilde{H}.)$ is in $\mathfrak{G}_0^S$ and Γ'_* is invertible in $\vec{\mathfrak{G}}_0^S$. Thus the path (4.18) is equal to

$$(4.20)\qquad (\tilde{C}.) \xrightarrow{\Gamma_\circ} (\tilde{H}.) \xleftarrow{\Delta_\circ} (\tilde{G}.)$$

with $\Gamma_\circ = \Delta_* \tilde{\Gamma}$, $\Delta_\circ = \Gamma'_* \tilde{\Delta}'$. It follows that the path (4.16) is equal in $\vec{\mathfrak{G}}_0^S$ to

$$(4.21)\qquad (A;C.) \xleftarrow{\Pi} (Y;\tilde{C}.) \xrightarrow{\Gamma_\circ} (Y;\tilde{H}.) \xleftarrow{\Delta_\circ} (Y;\tilde{G}.) \xrightarrow{\Pi} (X;G.)\,.$$

Now let

$$(4.22)\qquad (A;C.) \xleftarrow{\Pi} (AX;C^*.) \xrightarrow{\Gamma^*} (AX;K.) \xleftarrow{\Delta^*} (AX;G^*.) \xrightarrow{\Pi} (X;G.)$$

be the standard $\psi\varphi$-path from $(A;C.)$ to $(X;G.)$. This path is obviously equal to the 'blow-up'

$$(4.23)\qquad (A;C.) \xleftarrow{\Pi} (Y;\tilde{C}.) \xrightarrow{\tilde{\Gamma}^*} (Y;\tilde{K}.) \xleftarrow{\tilde{\Delta}^*} (Y;\tilde{G}.) \xrightarrow{\Pi} (X;G.)\ .$$

* We use various self-evident notations, in (4.16) and the rest of the proof, to indicate the effect of blowing up the indexing set for a direct system.

However, a further application of Lemma 4.4 tells us that the morphisms (4.23) and (4.21) are equal, so the theorem is completely proved.

Let $\mathfrak{G}_0^{SP}$ be the subcategory of $\vec{\mathfrak{G}}_0^S$ consisting of standard paths. The following corollary expresses the essential content of Theorem 4.15.

Corollary 4.24. *Let* $L : \vec{\mathfrak{G}}_0^S \to \mathfrak{G}_1$ *be the natural extension of* $\varinjlim$ *and let* $L' = L \mid \mathfrak{G}_0^{SP}$. *Then*

$$L' : \mathfrak{G}_0^{SP} \to \mathfrak{G}_1$$

is full and faithful and surjective on objects.

A very minor modification of the argument of Theorem 3.18 yields

Theorem 4.25. *Replace Hypothesis 3.1 with Hypothesis 4.1 in the enunciation of Theorem 3.18. Then the conclusion holds except that (3.19) is only asserted for morphisms of* $\mathfrak{G}_0^S$ *which are the identity on the indexing ordered sets.*

Proof. We have the diagram (in view of Proposition 3.17 and Corollary 4.24)

$$\begin{tikzcd}
 & \mathfrak{G}_0^S \arrow[r, "F_0^S"] \arrow[d, "i"] & \mathfrak{C}^S \arrow[d, "i"] \\
\mathfrak{G}_0^{SP} \arrow[r, "j"] \arrow[dr, "L'"'] & \vec{\mathfrak{G}}_0^S \arrow[r, "\vec{F}_0^S"] \arrow[d, "L"] & \vec{\mathfrak{C}}^S \arrow[d, "L"] \\
 & \mathfrak{G}_1 \arrow[r, "F_1"] & \mathfrak{C}
\end{tikzcd}$$

We have $F_1 L' = L\vec{F}_0^S j$, but we do not guarantee $F_1 L = L\vec{F}_0^S$. Thus F_1 satisfies (3.19) on those morphisms of $\mathfrak{G}_0^S$ whose i-image lies in $\mathfrak{G}_0^{SP}$. It is easily shown that any morphism of $\mathfrak{G}_0^S$ which is the identity on the indexing set is a standard path *, so the conclusion follows; the uniqueness of F_1 is based, of course, on Lemma 4.4 which shows that every φ is the limit of a suitable morphism of $\mathfrak{G}_0^S$, namely Γ.

Remark. We could certainly weaken the hypotheses of Theorem 4.25, since we only need F_0^S to be defined on $\mathfrak{G}_0^{SP}$; however, such a refinement of the enunciation would be premature since we do not know if $\mathfrak{G}_0^{SP}$ is strictly smaller than $\vec{\mathfrak{G}}_0^S$. It certainly coincides with $\mathfrak{G}_0^S$ if both Hypotheses 3.1 and 4.1 are satisfied.

* The argument is a very slight generalization of that proving part (ii) of Theorem 4.15.

§ 5. Direct limits and exactness

There does not appear to be a categorical notion of exactness in $\mathfrak{C}^S$ even where there is a categorical notion of exactness in $\mathfrak{C}$. However we may introduce the notion of *local exactness* in $\mathfrak{C}^S$ if exactness is defined in $\mathfrak{C}$, namely,

$$(5.1) \qquad (A;C.) \xrightarrow{\Phi} (B;D.) \xrightarrow{\Psi} (X;E.) \quad ,$$

is *locally exact* if, for every $\alpha \in A$,

$$C_\alpha \xrightarrow{\varphi_\alpha} D_{\rho(\alpha)} \xrightarrow{\psi_{\rho(\alpha)}} E_{\sigma\rho(\alpha)}$$

is exact; here, as usual, $\Phi = (\rho;\varphi.)$, $\Psi = (\sigma;\psi.)$, and

$$C. = \{C_\alpha; f_{\alpha\alpha'}\}\ , \qquad D. = \{D_\beta; g_{\beta\beta'}\}\ , \qquad E. = \{E_\rho; h_{\rho\rho'}\}\ .$$

We remark that if Φ, Ψ in (5.1) belong to $\mathfrak{C}^A$, then (5.1) is locally exact if and only if it is exact in $\mathfrak{C}^A$. We prove a theorem which serves to justify our insistence that the functions ρ, σ be cofinal.

Theorem 5.2. *Let (5.1) be a locally exact sequence in* $\mathfrak{G}^S$. *Then the limit sequence*

$$(5.3) \qquad C \xrightarrow{\varphi} D \xrightarrow{\psi} E$$

is exact.

Proof. Certainly $\ker\psi \supseteq \operatorname{im}\varphi$. Now let $y \in D$ with $\psi(y) = e$ (where e is as usual, the identity element in the appropriate group). Then $y = [y_\beta]$ for some $\beta \in B$, $y_\beta \in D_\beta$ and $[\psi_\beta(y_\beta)] = e$. Thus there exists $\xi \geqslant \sigma(\beta)$ such that $h_{\sigma(\beta),\xi}\,\psi_\beta(y_\beta) = e$. Choose $\beta' \geqslant \beta$ with $\sigma(\beta') \geqslant \xi$. Then

$$\psi_{\beta'} g_{\beta\beta'}(y_\beta) = h_{\sigma(\beta),\sigma(\beta')}\,\psi_\beta(y_\beta) = e\ .$$

Set $y_{\beta'} = g_{\beta\beta'}(y_\beta)$. Then $y = [y_{\beta'}]$ and $\psi_{\beta'}(y_{\beta'}) = e$.

Now choose α so that $\rho(\alpha) \geqslant \beta'$, and set $y_{\rho(\alpha)} = g_{\beta',\rho(\alpha)} y_{\beta'}$. Then $y = [y_{\rho(\alpha)}]$ and $\psi_{\rho(\alpha)}(y_{\rho(\alpha)}) = e$. By virtue of the local exactness of (5.1) we have $y_{\rho(\alpha)} = \varphi_\alpha(x_\alpha)$ for some $x_\alpha \in C_\alpha$. Thus $y = \varphi[x_\alpha]$ and the theorem is proved.

Of course, Theorem 5.2 carries the implication that locally surjective (injective) morphisms of $\mathfrak{G}^S$ pass in the limit to surjective (injective) homomorphisms.

§6. Fibre-maps

The notion of fibre-map given in §2 is perfectly adequate for our purpose. However, a stronger notion would seem to be necessary if one is to develop a 'homotopy theory' for the category $\mathfrak{D}\mathfrak{S}$.

Definition 6.1. Let $\rho : A \to B$ be a morphism of $\mathfrak{D}\mathfrak{S}$ Let $\bar{A} \subseteq A \times B$ be the set of all pairs (α, β') with $\beta' \geqslant \rho(\alpha)$. A *strong fibre-map* from A to B is a pair (ρ, θ) where $\rho : A \to B$ and θ is a monotone function $\bar{A} \to A$ with $\theta(\alpha, \rho(\alpha)) = \alpha$, $\theta(\alpha, \beta') \geqslant \alpha$ and $\rho\theta(\alpha, \beta') = \beta'$.

We may loosely say that ρ is itself a strong fibre-map; then it is plain that a strong fibre-map is a fibre-map. It is also plain that the function $\bar{\rho} : \bar{A} \to B$ constructed in (2.3) is a strong fibre-map, with $\theta((\alpha, \beta), \beta') = (\alpha, \beta')$. We may also append to Theorem 2.8 the complement.

Theorem 6.2. *If σ in (2.6) is a strong fibre-map, so is ρ'.*

Proof. Let $\theta : \bar{B} \to B$ be given so that (σ, θ) is a strong fibre-map from B to X. We define $\theta' : \bar{Y} \to Y$ by

$$\theta'((\alpha, \beta), \alpha') = (\alpha', \theta(\beta, \rho(\alpha'))) , \quad \rho(\alpha) = \sigma(\beta) , \quad \alpha' \geqslant \alpha .$$

Plainly the composite of two fibre-maps is a fibre-map. Likewise we have

Theorem 6.3. *The composite of two strong fibre-maps is a strong fibre-map.*

Proof. Suppose given $\rho : A \to B$ with $\theta_1 : \bar{A} \to A$ and $\sigma : B \to C$ with $\theta_2 : \bar{B} \to B$. Let $\bar{\bar{A}} \subseteq A \times C$ consist of pairs (α, γ) with $\gamma \geqslant \sigma\rho(\alpha)$. We define $\theta : \bar{\bar{A}} \to A$ by

$$\theta(\alpha, \gamma) = \theta_1(\alpha, \theta_2(\rho(\alpha), \gamma)) .$$

Then θ is well-defined since $\gamma \geqslant \sigma\rho(\alpha)$ and $\theta_2(\rho(\alpha), \gamma) \geqslant \rho(\alpha)$ and it is easy to see that $(\sigma\rho, \theta)$ is a strong fibre-map. Composition of strong fibre-maps is also plainly associative in the obvious sense.

We give one, rather obvious, application of the notion of a strong fibre-map. Let $\rho, \rho' : A \to B$; we then write $\rho \leqslant \rho'$ if $\rho(\alpha) \leqslant \rho'(\alpha)$ for every $\alpha \in A$. We note that if $\Phi = (\rho; \varphi_\cdot) : (A; C_\cdot) \to (B; D_\cdot)$ in $\mathfrak{C}^S$, and if $\rho \leqslant \rho'$, then there is a well-defined morphism $\Phi' = (\rho'; \varphi'_\cdot) : (A; C_\cdot) \to (B; D_\cdot)$ where $\varphi'_\cdot$ is given by

$$\varphi'_\alpha = g_{\rho(\alpha), \rho'(\alpha)} \varphi_\alpha ,$$

and that, if $\mathfrak{C}$ is cocomplete, then Φ and Φ' induce the same morphism of direct limits. We prove

Theorem 6.4. *Suppose given a strong fibre-map $(\rho, \theta) : A \to B$, a function $\tau : X \to A$ in $\mathfrak{D}\mathfrak{S}$ and functions $\sigma \leqslant \sigma' : X \to B$ in $\mathfrak{D}\mathfrak{S}$ with $\sigma = \rho\tau$. Then we may lift the inequality to an inequality $\tau \leqslant \tau' : X \to A$ with $\sigma' = \rho\tau'$. Moreover $\tau'(\xi) = \tau(\xi)$ if $\sigma'(\xi) = \sigma(\xi)$.*

Proof. We simply set

$$\tau'(\xi) = \theta(\tau(\xi), \sigma'(\xi)) .$$

Plainly τ' is monotone and $\tau' \geqslant \tau$, so that τ' is certainly cofinal. Also $\rho\tau'(\xi) = \sigma'(\xi)$ and $\tau'(\xi) = \tau(\xi)$ if $\sigma'(\xi) = \sigma(\xi)$ $(= \rho\tau(\xi))$.

Theorem 6.4 is a candidate to be a 'homotopy lifting theorem' in any reasonable 'homotopy theory' in $\mathfrak{D}\mathfrak{S}$.

§ 7. The relation to the Kan extension

The author has learnt, in conversation with Fritz Ulmer, of a very general and elegant approach to the same problem which also handles the special case with which we are concerned in [6]. Consider the diagram of categories and functors

(7.1)
$$\begin{tikzcd}
(\mathfrak{H}^{\mathrm{opp}}, \mathfrak{S}) \arrow[ddr, bend left, "E_Q(F)"] & \\
\mathfrak{Y} \arrow[u, "R"'] \arrow[dr, "E_J(F)"] & \\
\mathfrak{H} \arrow[u, "J"'] \arrow[uu, bend left, "Q"] \arrow[r, "F"'] & \mathfrak{C}
\end{tikzcd}$$

where Q is the Yoneda embedding, R is the regular representation $R(Y) = \mathfrak{Y}(J-, Y)$, F is additive, and $E_J(F)$, $E_Q(F)$ are Kan extensions. Then Ulmer remarks that if F is additive (i) $E_Q(F) \circ R = E_J(F)$, (ii) $E_Q(F)$ has a right adjoint. It follows that $E_J(F)$ preserves all colimits preserved by R. Now R preserves $\varinjlim_\alpha Y_\alpha$ if, for all objects X in $\mathfrak{H}$,

(7.2)
$$\mathfrak{Y}(JX, \varinjlim_\alpha Y_\alpha) = \varinjlim_\alpha \mathfrak{Y}(JX, Y_\alpha) .$$

If we take the example of a full embedding $J : \mathfrak{G}_0 \subseteq \mathfrak{G}_1$ of one category of groups in another, then we are asking if

(7.3)
$$\operatorname{Hom}(G_0, \varinjlim_\alpha G_\alpha) = \varinjlim_\alpha \operatorname{Hom}(G_0, G_\alpha) , \qquad G_0 \in |\mathfrak{G}_0| .$$

Let us suppose that, as in our case, $\mathfrak{G}_1$ consists precisely of direct limits of direct systems in $\mathfrak{G}_0$ and let us demand (7.3) if $G_\alpha \in |\mathfrak{G}_0|$. Then we may conclude that, if $G = \varinjlim_\alpha G_\alpha$,

$$E_J(F)(G) = E_J(F)(\varinjlim_\alpha G_\alpha) = \varinjlim_\alpha E_J(F)(G_\alpha) = \varinjlim_\alpha F(G_\alpha) ,$$

and the Kan extension yields the functor F_1 of Theorem 3.18. Thus, if F is additive, condition (7.3) produces the conclusion of Theorem 3.18 without requiring Hypothesis 3.1 and the special property of $F = F_0$ in the statement of that theorem. Indeed the special property referred to appears as a consequence of condition (7.3). On the other hand there will be cases in which Hypothesis 3.1 is satisfied and condition (7.3) is false, e.g., if $\mathfrak{G}_0$ is the category of countable Abelian groups. We remark, in particular, that condition (7.3) does hold if we take $\mathfrak{G}_0$ to be the category of finitely-generated Abelian groups which is, as we mentioned in the Introduction, the case considered in [6].

The author is very much indebted to Fritz Ulmer for clarifying the relation of the procedure described in this paper to the more general approach via the Kan extension outlined above. A further study of this connection will form the subject-matter of a subsequent paper.

References

[1] M.Artin and A.Grothendieck, Cohomologie étale des schemas, Sém. de Géom. Alg . I.H.E.S. (1963-1964).
[2] B.Eckmann and P.J.Hilton, Commuting limits with colimits, J. Alg. 11 (1969) 116–144.
[3] P.Gabriel and M.Zisman, Calculus of fractions and homotopy theory (Springer-Verlag, 1967).
[4] J.W.Gray, Fibred and cofibred categories, Proc. La Jolla Conf. on Categorical Algebra (Springer-Verlag, 1966).
[5] A.Grothendieck, Catégories fibrées et descente, Sém. de Géom. Alg., I.H.E.S. (1961).
[6] P.J.Hilton, Putting coefficients into a cohomology theory, Indag. Math. (1970).

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 1 (1971) pp. 27–40.

H-SPACES WHICH ARE BUNDLES OVER S^7

Morton CURTIS * and Guido MISLIN **
Rice University and University of California, Berkeley

Received February 1970

Introduction

It is the purpose of this paper to prove the following results.

Theorem I. *If the total space X of an* SO(4) *3-sphere bundle over S^7 is an H-space, then X has the homotopy type of the total space of a principal S^3 bundle over S^7.*

Hence such an H-space has the homotopy type of one of the following spaces:
(1) $S^3 \times S^7$ or Sp(2);
(2) the "Hilton-Roitberg criminal" [2] and [3];
(3) one of the two new H-spaces of Stasheff [7];
(4) one of the two remaining homotopy types of total spaces of principal S^3 bundles over S^7.

It is not known, if the spaces of (4) are actually H-spaces. We will prove that one of them is if and only if the other one is.

Theorem II. (i) *The total space of all principal* SU(3) *bundles over S^7 are H-spaces.*

(ii) *There are exactly four homotopy types of these total spaces.*

The known H-spaces SU(4) and $S^7 \times$ SU(3) are of two of these and the other two are new ones. They have of course the homotopy type of fifteen dimensional differentiable manifolds.

Section 1 is devoted to some well-known elementary properties of H-spaces. The main result is Theorem 1.3 which tells when the pull-back of H-spaces and H-maps yields an H-space. Section 2 outlines the method of Zabrodsky [9] of mixing homotopy types. It winds up with a convenient criterion (Theorem 2.5) for a space to be

* Research partially supported by NSF Grant No. GP 12715.
** Research supported by the Schweizerischer Nationalfonds.

an H-space. In section 3 we prove Theorem I. Results of the first two sections are applied to determine which SO(4) 3-sphere bundles over S^7 are H-spaces. We recover the results of Stasheff [7] for those bundles whose structure group reduces to $SU(2) = S^3$, and show that the remaining three homotopy types of SO(4) 3-sphere bundles are not H-spaces. The results of James and Whitehead [4] and [5] are used to count the homotopy types here.

Finally in section 4 we prove Theorem II (i) using the results of the first two sections; part (ii) is shown to follow immediately by looking at the homotopy groups of the spaces involved.

The authors wish to thank Emery Thomas for helping to simplify the proof of Theorem II, and to George Cooke for many stimulating conversations on the subject matter of this paper.

§ 1. Elementary homotopy properties of H-spaces

All spaces are supposed to have the homotopy type of a 1-connected CW complex with finite skeleta. Furthermore, all our spaces have (nice) basepoints, which we suppose to be preserved by maps and homotopies. By an H-space we mean a space X together with a map $m : X \times X \to X$ which is, when restricted to the wedge, homotopic to the folding map $\nabla : X \vee X \to X$. (Of course this means that the basepoint of X acts as a two-sided homotopy identity via m.) If $m : X \times X \to X$ and $n : Y \times Y \to Y$ are H-spaces then $f : X \to Y$ is called an H-map if $n(f \times f) \simeq fm$. We remark that this implies that a map which is homotopic to an H-map is actually an H-map.

We will need to know that in certain circumstances the pull-back of a diagram of H-spaces and H-maps is an H-space. For this we prove the following "pullback lemma".

Lemma 1.1. *Consider the diagram*

$$\begin{tikzcd}
A \arrow[r, "i"] & X \arrow[dr, "h"] \arrow[drr, "g"] \arrow[ddr, "f"'] & & \\
& & D \arrow[r, "q"] \arrow[d, "p"] & C \arrow[d, "\psi"] \\
& & A \arrow[r, "\varphi"'] & B
\end{tikzcd}$$

where 1) *D is the pullback of φ and ψ; $D = \{(a, c) \mid \varphi a = \psi c\} \subset A \times C$;*

2) *ψ is a fibration;*

3) *i is a cofibration;*

4) *$\varphi fi = \psi gi$ and $\varphi f \simeq \psi g$.*

Then one can construct a map $h : X \to D$ such that the whole diagram is homotopy commutative and $fi = phi$, $gi = qhi$.

Proof. Define $h_A : A \to D$ by the universal property of the pullback: $h_A(a) = (fia, gia)$. Now we extend h_A to X in the following way. By assumption $\varphi fi = \psi fi$ and $\varphi f \simeq \psi g$. Since ψ is a fibration, we can lift this homotopy to get a map $g' : X \to C$ with $\psi g' = \varphi f$. But (X, A) being a cofiber pair we can, according to the relative covering homotopy theorem, find $g' : X \to C$ such that $gi = g'i$. Defining $h : X \to D$ by $h(x) = (fx, g'x)$ our assertion follows.

The next lemma is trivial.

Lemma 1.2. *Let X be an H-space with structure map $m : X \times X \to X$. Then there exists an H-space structure $m' : X \times X \to X$ homotopic to m such that $(m' | X \vee X) = \nabla : X \vee X \to X$, that is m' has a strict identity. Furthermore, if $f : X \to Y$ is an H-map and we replace the structure maps of X and Y by homotopic ones, then f is still an H-map.*

For the first part, extend the homotopy $\nabla \simeq (m \mid X \vee X)$ to $X \times X$ to get m'. This can be done since our spaces have nice basepoints. The other statement is obvious.

Theorem 1.3. *In the diagram*

$$\begin{tikzcd}
Y \arrow[r, "f'"] \arrow[d, "p'"] & E \arrow[d, "p"] \\
X \arrow[r, "f"] & B
\end{tikzcd}$$

let X, E and B be H-spaces, p and f H-maps and Y their pullback. If p is a fibration, then Y admits an H-space structure such that f' and p' are H-maps.

Proof. We may assume by the previous lemma that X, E and B are H-spaces with strict identities. We have the following diagram

$$\begin{tikzcd}
Y \vee Y \arrow[r, "i"] & Y \times Y \arrow[rrr, "f' \times f'"] \arrow[dr, dashed, "\mu"] \arrow[ddd, "p' \times p'"'] & & & E \times E \arrow[dl, "\gamma"'] \arrow[ddd, "p \times p"] \\
& & Y \arrow[r, "f'"] \arrow[d, "p'"] & E \arrow[d, "p"] & \\
& & X \arrow[r, "f"] & B & \\
& X \times X \arrow[ur, "\alpha"] \arrow[rrr, "f \times f"] & & & B \times B \arrow[ul, "\beta"']
\end{tikzcd}$$

Since α and γ are H-space structures with strict identities it follows that $f\alpha(p' \times p')\, i = p\gamma(f' \times f')\, i$ and, f and p being H-maps, we have $f\alpha(p' \times p') \simeq p\gamma(f' \times f')$. Using the "pullback lemma" (1.1) we construct μ in a way that $p'\mu i = \alpha(p' \times p')i$ and $f'\mu i = \gamma(f' \times f')\, i$. But these two maps agree with $p'\nabla$ respectively $f'\nabla$ since α and γ have strict identities. It follows from the universal property of the pullback that $\mu i = \nabla : Y \vee Y \to Y$. Hence μ is an H-space structure for Y and, by the homotopy commutativity of the corresponding quadrangles, we see that f' and p' are H-maps.

Definition 1.4. If $m : X \times X \to X$ is an H-space then $c \in H^*(X;G)$ is called primitive if

$$m^*c = p_1^*c + p_2^*c$$

where p_1 and p_2 are the projections.

Theorem 1.5. *Let X be an H-space and $K(G, n)$ an Eilenberg-MacLane space with its canonical H-space structure. If $i \in H^n(K(G, n); G)$ is a characteristic element, then $f: X \to K(G, n)$ is an H-map if and only if $f^*i \in H^n(X; G)$ is primitive.*

Proof. Consider the diagram

$$\begin{tikzcd} X \times X \arrow[r, "f \times f"] \arrow[d, "p_1, p_2, \mu"'] & K \times K \arrow[d, "\pi_1, \pi_2, m"] \\ X \arrow[r, "f"] & K \end{tikzcd}$$

where μ and m are the multiplications and p_i, π_i the projections.

(1) $\qquad fp_i = \pi_i(f \times f)$ and, since i is primitive,

(2) $\qquad m^*i = \pi_1^*i + \pi_2^*i$.

The condition that f be an H-map means

$$f\mu \simeq m(f \times f),$$

and since we are mapping into an Eilenberg-MacLane space, this is equivalent to

$$f^*i = (f \times f)^* m^*i \ , \quad \text{or}$$

$$f^*i = (f \times f)^* (\pi_1^*i + \pi_2^*i) \ , \quad \text{by (2)}$$

$$= p_1^*f^*i + p_2^*f^*i \ , \qquad \text{by (1)}.$$

This latter is just the condition that f^*i be primitive.

§ 2. Mixing homotopy types

In this section we outline the technique of mixing homotopy types due to Zabrodsky [9]. This is the key tool for the results of sections 3 and 4. A map $f: X \to Y$ is a *rational equivalence* if

$$f^*: H^*(Y;Q) \to H^*(X;Q)$$

is an isomorphism. It is a *q-equivalence* if it induces an isomorphism with $\mathbf{Z}_q$ replacing Q. Of course a q-equivalence is a q^k-equivalence for all $k \geq 1$. If $\mathcal{P}_1$ is a set of primes, a *$\mathcal{P}_1$-equivalence* means a p-equivalence for each $p \in \mathcal{P}_1$. Denote by $\mathcal{P}$ the set of all primes. Then a $\mathcal{P}$-equivalence is a homotopy equivalence, since our spaces are "nice".

Lemma 2.1 (Zabrodsky). (i) *If $f: X \to Y$ is a rational equivalence and $\mathcal{P}_1$ a set of primes, then f may be factored as*

$$X \xrightarrow{f_1} X(\mathcal{P}_1) \xrightarrow{f_2} Y$$

where f_1 is a $\mathcal{P}_1$-equivalence and f_2 a fibration which is a $(\mathcal{P} - \mathcal{P}_1)$-equivalence.

(ii) *If, in addition, X and Y are H-spaces and if f is an H-map, then $X(\mathcal{P}_1)$ is an H-space and f_1 and f_2 are H-maps.*

The proof of part (i) uses a Moore-Postnikov system and is straightforward. For part (ii) one has to see that one can put an H-space structure on the spaces in the Postnikov tower such that all maps are H-maps.

By a factorization of a rational equivalence with respect to a set of primes we will always mean a factorization as described in Lemma 2.1. To prove the main result of this section, Theorem 2.5, we need a kind of uniqueness result for factorizations, and this is furnished by the next lemma.

Lemma 2.2. *Given factorizations of two rational equivalences $X \xrightarrow{f} Z \xleftarrow{g} Y$ with respect to $\mathcal{P}_1$, and a $\mathcal{P}_1$-equivalence $\varphi: X \to Y$. Then there exist homotopy equivalences α, β making the triangles below commutative.*

$$\begin{tikzcd}
X \arrow[r, "f_1"] \arrow[d, "\varphi"'] & X(\mathcal{P}_1) \arrow[d, dashed, shift left, "\beta"] \arrow[dr, "f_2"] & \\
 & & Z \\
Y \arrow[r, "g_1"'] & Y(\mathcal{P}_1) \arrow[u, dashed, shift left, "\alpha"] \arrow[ur, "g_2"'] &
\end{tikzcd}$$

To construct α consider the diagram

$$
\begin{tikzcd}
X \arrow[rr, "f_1"] \arrow[d, "\varphi"'] & & X(\mathcal{P}_1) \arrow[dd, "f_2"] \\
Y \arrow[d, "g_1"'] & & \\
Y(\mathcal{P}_1) \arrow[rr] \arrow[uurr, dashed, "\alpha"] & & Z
\end{tikzcd}
$$

and note that $g_1\varphi$ is a $\mathcal{P}_1$-equivalence. The obstruction to constructing α lies in the cohomology of $Y(\mathcal{P}_1)$ mod X, with coefficients in the homotopy groups of the fiber of the fibration f_2. Since f_2 is a p^k-equivalence for all $k \geq 1$, these coefficient groups are such that

$$
H^*(Y(\mathcal{P}_1)) \xrightarrow{(g_1\varphi)^*} H^*(X)
$$

is an isomorphism. Thus the obstruction vanishes. β may be constructed similarly.

Lemma 2.3. *Let A be the pullback in the following diagram in which f is a fibration.*

$$
\begin{tikzcd}
A \arrow[r, "h"] \arrow[d, "j"'] & B \arrow[d, "f"] \\
C \arrow[r, "g"] & D
\end{tikzcd}
$$

If g is a $\mathcal{P}_1$-equivalence, then so is h.

We simply note that since f is a fibration, h and g have the same "fiber".

Lemma 2.4. *Given two factorizations of the rational equivalence $f : X \to Y$ with respect to sets of primes $\mathcal{P}_1$ and $\mathcal{P}_2$ such that $\mathcal{P}_1 \cup \mathcal{P}_2 = \mathcal{P}$. Then X has the homotopy type of the pullback of the fibrations f_2 and f_2'.*

$$
\begin{tikzcd}
 & & X(\mathcal{P}_1) \arrow[dr, "f_2"] & \\
X \arrow[r, "\varphi"] \arrow[urr, "f_1"] \arrow[drr, "f_1'"'] & P \arrow[ur, "p_1"'] \arrow[dr, "p_2"] & & Y \\
 & & X(\mathcal{P}_2) \arrow[ur, "f_2'"'] &
\end{tikzcd}
$$

Let P be the pullback of f_2 and f_2'. Then P will have the homotopy type of a 1-connected CW complex with finite skeleta, since X and Y are spaces of this kind. By 2.3, p_i is a $\mathcal{P}_i$-equivalence for $i = 1, 2$. Since also f_i is a $\mathcal{P}_i$-equivalence for $i = 1, 2$, it follows that the canonical map $\varphi : X \to P$ is a $(\mathcal{P}_1 \cup \mathcal{P}_2)$-equivalence, which means it is a homotopy equivalence since $\mathcal{P}_1 \cup \mathcal{P}_2 = \mathcal{P}$.

Theorem 2.5. *Given $X_0 \xrightarrow{f} X_1 \xrightarrow{g} X_2$ where f is a $\mathcal{P}_1$-equivalence and g is a $\mathcal{P}_2$-equivalence, $\mathcal{P}_1 \cup \mathcal{P}_2 = \mathcal{P}$. Suppose X_0 and X_2 are H-spaces of the homotopy type of finite CW complexes such that $H^*(X_2;\mathbb{Q}) = \Lambda(x_1, \ldots, x_n)$ is primitively generated. Further suppose that*

$$f^*g^* : H^*(X_2;\mathbb{Q}) \longrightarrow H^*(X_0;\mathbb{Q})$$

maps primitives to primitives. Then X_1 is an H-space with the homotopy type of a finite CW complex.

Proof. Since $H^*(X_2;\mathbb{Q})$ is primitively generated we can find (using 1.5) a rational equivalence

$$\varphi : X_2 \longrightarrow K = \Pi\, K(\mathbb{Z}; n_i) \quad (n_i = \text{degree } x_i)$$

such that φ is an H-map. We can even choose φ in such a way that in addition $\varphi g f$ is an H-map, since f^*g^* maps primitives to primitives. By 2.1 we have the following factorization, with h_0, h_2 and k_0, k_2 H-maps.

$$\begin{tikzcd}
 & & & X_0(\mathcal{P}_1) \arrow[dddr, "h_0"] & \\
 & & & X_1(\mathcal{P}_1) \arrow[u, dashed, "\beta"] \arrow[ddr, "h_1"] & \\
 & & & X_2(\mathcal{P}_1) \arrow[dr, "h_2"] & \\
X_0 \arrow[r, "f"] \arrow[uuurrr] \arrow[dddrrr] & X_1 \arrow[r, "g"] \arrow[uurr] \arrow[ddrr] & X_2 \arrow[ur] \arrow[rr, "\varphi"] \arrow[dr] & & K \\
 & & & X_2(\mathcal{P}_2) \arrow[ur, "k_2"] & \\
 & & & X_1(\mathcal{P}_2) \arrow[u, dashed, "\alpha"] \arrow[uur, "k_1"] & \\
 & & & X_0(\mathcal{P}_2) \arrow[uuur, "k_0"] &
\end{tikzcd}$$

By 2.2 we have maps α and β as shown and they are homotopy equivalences. Thus X_1 has the homotopy type of the pullback of

$$\begin{tikzcd} & X_0(\mathcal{P}_1) \arrow[d, "h_0"] \\ X_2(\mathcal{P}_2) \arrow[r, "k_2"] & K \end{tikzcd}$$

by 2.4. Hence X_1 is an H-space. Since X_1 is p-equivalent to X_0 respectively X_2, its cohomology $H^*(X_1; \mathbf{Z})$ is of finite type. By hypothesis X_1 is simply connected, so it has the homotopy type of a finite CW complex.

§ 3. 3-sphere bundles over S^7

An S^3 bundle over S^7 with structure group SO(4) is classified up to bundle isomorphism by its characteristic element $\chi \in \pi_6(\mathrm{SO}(4))$. For the homotopy classification of the total spaces of such bundles the following diagram is crucial:

$$\pi_9(S^6) \xrightarrow{(p_*\chi)_*} \pi_9(S^3) \xleftarrow{\ J\ } \pi_6(\mathrm{SO}(3)) \xrightarrow{\ i_*\ } \pi_6(\mathrm{SO}(4)) \xrightarrow{\ p_*\ } \pi_6(S^3)\,.$$

Here i_* and p_* are induced by the natural maps $\mathrm{SO}(3) \to \mathrm{SO}(4) \to S^3$ and $(p_*\chi)_*$ denotes composition with $p_*\chi \in \pi_6(S^3)$. Note that one has a section $s : S^3 \to \mathrm{SO}(4)$, which yields complete information about the maps i_* and p_*. Following James and Whitehead [5] one has the following theorem.

Theorem 3.1. *Let B_1 and B_2 be the total spaces of two* SO(4) *3-sphere bundles over S^7 classified by χ_1 and χ_2 in $\pi_6(\mathrm{SO}(4))$. Denote by $G(\chi_i)$ the subgroup $i_*J^{-1}(p_*\chi_i)_*\pi_9(S^6)$ of $\pi_6(\mathrm{SO}(4))$ and let $G_i = \{G(\chi_i)\chi_i \cup G(\chi_i)(-\chi_i)\} \subset \pi_6(\mathrm{SO}(4))$ for $i = 1, 2$. Then $B_1 \simeq B_2$ if and only if $p_*\chi_2 = \pm p_*\chi_2$ and $G_1 = G_2$.*

Corollary 3.2. *There are exactly* 10 *different homotopy types of total spaces of* SO(4) *3-sphere bundles over S^7.*

Proof. This follows by checking the conditions given in Theorem 3.1 using that

(1) $\quad \alpha_* : \pi_9(S^6) \cong \mathbf{Z}_{24} \to \pi_9(S^3) \cong \mathbf{Z}_3$ is surjective for a generator $\alpha \in \pi_6(S^3)$;

(2) $\quad J : \pi_6(\mathrm{SO}(3)) \cong \mathbf{Z}_{12} \to \pi_9(S^3) \cong \mathbf{Z}_3$ is surjective.

Part (1) follows immediately from the fact that $\alpha \circ \Sigma^3\alpha$ is a generator of $\pi_9(S^3)$ by [6]. Since S^3 is an H-space we further conclude that $(n\alpha)_*\beta = (n\alpha) \circ \beta = n(\alpha \circ \beta) = n(\alpha_*\beta)$ by applying the left-distributive law. Hence $(n\alpha)_* = 0$ if and only if $n \equiv 0\ (3)$. To prove (2) one uses that $J : \pi_3(\mathrm{SO}(3)) \to \pi_6(S^3)$ is onto [8], and that the following diagram commutes by naturality:

$$\begin{array}{ccc} \pi_3(\mathrm{SO}(3)) & \xrightarrow{\ \alpha^*\ } & \pi_6(\mathrm{SO}(3)) \\ \downarrow{\scriptstyle J} & & \downarrow{\scriptstyle J} \\ \pi_6(S^3) & \xrightarrow{(\Sigma^3\alpha)^*} & \pi_9(S^3) \end{array}$$

where $(\Sigma^3\alpha)^*$ is onto since $(\Sigma^3\alpha)^*\alpha = \alpha \circ \Sigma^3\alpha$. It follows that one gets exactly one, respectively two, homotopy types with the same $\pm p_*\chi$ according to whether $(p_*\chi)_* : \pi_9(S^6) \to \pi_9(S^3)$ is surjective or not.

We will denote these homotopy types by

$$\begin{array}{ccccccc} X_{0,0} & X_{1,0} & X_{2,0} & X_{3,0} & X_{4,0} & X_{5,0} & X_{6,0} \\ X_{0,1} & & & X_{3,1} & & & X_{6,1} \end{array}$$

with $p_*\chi(X_{m,0}) = \pm p_*\chi(X_{m,1}) = \pm m\alpha$, where we fix the generator α by $p_*\chi(\mathrm{Sp}(2)) = \alpha$, i.e., $X_{1,0} \simeq \mathrm{Sp}(2)$. Further we fix the second index in $X_{m,n}$ by the condition that $X_{m,0}$ has the homotopy type of the total space of a principal S^3 bundle over S^7. This is compatible with our notation, since there are exactly 7 such homotopy types according to [3]. Hence we have $X_{0,0} \simeq S^3 \times S^7$, $X_{1,0} \simeq \mathrm{Sp}(2)$ and so on. Note that $X_{0,1}$ has the homotopy type of a bundle with a cross section, since $p_*\chi(X_{0,1}) = 0$. It is convenient to introduce the following convention:

We define $X_{m,n}$ for all $(m, n) \in \mathbf{Z}_{12} \times \mathbf{Z}_3$ by $X_{m,n} = X_{m,-n} = X_{-m,n} = X_{-m,-n}$. As an obvious consequence we have

Lemma 3.3. *If $B \simeq X_{m,n}$ is an SO(4) 3-sphere bundle over S^7 classified by $\chi \in \pi_6(\mathrm{SO}(4))$ and B' is another one classified by $r\chi$, then $B' \simeq X_{rm,rn}$.*

Note that hence, if $p : B \to S^7$ denotes the bundle projection and $r : S^7 \to S^7$ a map of degree r, then the pullback of p and r has the homotopy type of $X_{rm,rn}$. We will write this pullback by abuse of notation in the following form:

$$\begin{tikzcd} X_{rm,rn} \arrow[r] \arrow[d] & X_{m,n} \arrow[d] \\ S^7 \arrow[r, "r"] & S^7 \end{tikzcd}$$

We are now ready to prove the Theorem I of the Introduction.

Theorem 3.4. * (i) (Stasheff [7]) *$X_{3,0}$, $X_{4,0}$, $X_{5,0}$ are H-spaces.*

(ii) *$X_{0,1}$, $X_{3,1}$, $X_{6,1}$ are not H-spaces.*

(iii) *$X_{2,0}$ is an H-space if and only if $X_{6,0}$ is.*

Proof. Consider the diagram

$$\begin{tikzcd} S^3 \times S^7 \simeq X_{12,0} \arrow[r, "\varphi"] \arrow[d] & X_{3,0} \arrow[r, "\psi"] \arrow[d] & X_{1,0} \simeq \mathrm{Sp}(2) \arrow[d] & \\ S^7 \arrow[r, "4"] & S^7 \arrow[r, "3"] & S^7 \arrow[r, "(1,0)"] & \mathrm{BSO}(4) \end{tikzcd},$$

* Parts (ii) and (iii) of this theorem have been obtained independently by Hilton and Roitberg.

where 4, 3 denote maps of the corresponding degrees. We can consider the two squares as pullbacks, by abuse of notation. By 2.3 φ is a $(\mathcal{P}\cdot\{2\})$-equivalence and ψ is a $(\mathcal{P}\cdot\{3\})$-equivalence. One sees easily that $H^*(X_{m,n};\mathbb{Q}) \cong \Lambda(x_3, x_7)$, which is of course primitively generated, if $X_{m,n}$ is an H-space. Moreover a rational equivalence $X_{m,n} \to X_{m',n'}$ pulls primitives back to primitives, since the primitives are exactly the elements of degree 3 respectively 7. So by 2.5 $X_{3,0}$ is an H-space. Interchanging 3 and 4 shows that $X_{4,0}$ is an H-space. If we let n denote the map on total spaces corresponding to a map of degree n on S^7, we have

$$X_{3,0} = X_{15,0} \xrightarrow{3} X_{5,0} \xrightarrow{5} X_{1,0} \simeq \mathrm{Sp}(2)$$

and, having proved that $X_{3,0}$ is an H-space, we have that $X_{5,0}$ is also.

The key tool used in (ii) is a theorem of James and Whitehead [4] which implies that $X_{0,1}$ as a non-product bundle with a cross section cannot be an H-space. Now if $X_{6,1}$ were an H-space, then the diagram $X_{0,0} \xrightarrow{3} X_{0,1} \xrightarrow{2} X_{6,1}$ would imply that $X_{0,1}$ is an H-space by 2.5. Thus $X_{6,1}$ cannot be an H-space. Looking at $X_{0,0} \xrightarrow{3} X_{0,1} \xrightarrow{4} X_{3,1}$ one sees that $X_{3,1}$ cannot be an H-space.

For part (iii) we note that if $X_{2,0}$ is an H-space, then the diagram $X_{0,0} \xrightarrow{2} X_{6,0} \xrightarrow{3} X_{2,0}$ shows that $X_{6,0}$ is an H-space. If $X_{6,0}$ is an H-space $X_{6,0} \xrightarrow{3} X_{2,0} \xrightarrow{2} X_{1,0}$ shows that $X_{2,0}$ is an H-space.

§4. Principal SU(3) bundles over S^7

In this section we prove Theorem II, namely that all principal SU(3) bundles over S^7 are H-spaces and that there are exactly four different homotopy types of such H-spaces. The technique is like that of section 3, using Lemma 2.3 and Theorem 2.5.

Let $SU(3) \to X \to S^7$ be a principal SU(3) bundle classified by $\beta : S^7 \to BSU(3)$. Since $\pi_7(BSU(3)) \cong \mathbf{Z}_6$, see [1], one has up to bundle isomorphism six different principal SU(3) bundles over S^7. If $\alpha \in \pi_7(BSU(3))$ classifies $SU(3) \to SU(4) \to S^7$, then α is a generator of $\pi_7(BSU(3))$. We denote by X_n the total space of the bundle classified by $n\alpha$. Then $X_0 \simeq SU(3) \times S^7$ and $X_1 \simeq SU(4)$. It is immediate that X_{-n} is homotopy equivalent to X_n. Hence one has at most four different homotopy types of such spaces, represented by X_0, X_1, X_2, X_3. To prove part (i) of Theorem II it suffices to prove the following.

Proposition 4.1. *The spaces X_2 and X_3 are H-spaces.*

Proof. We have a diagram

$$\begin{tikzcd} SU(3) \times S^7 \simeq X_6 \arrow[r, "\varphi"] \arrow[d] & X_2 \arrow[r, "\psi"] \arrow[d] & X_1 \simeq SU(4) \arrow[d] & \\ S^7 \arrow[r, "3"] & S^7 \arrow[r, "2"] & S^7 \arrow[r, "\alpha"] & BSU(3) \end{tikzcd}$$

where 3 and 2 denote maps of S^7 of those degrees. By 2.3, φ is a $(\mathcal{P}\text{-}\{3\})$-equivalence and ψ is a $(\mathcal{P}\text{-}\{2\})$-equivalence. Now $H^*(X_n;\mathbf{Q}) \cong \Lambda(x_3, x_5, x_7)$ with $\deg(x_i) = i$, and x_i is primitive if X_n is an H-space. Hence we may apply Theorem 2.5 to $X_6 \to X_2 \to X_1$ to conclude that X_2 is an H-space. Interchanging 2 and 3 shows X_3 is an H-space.

To see that the spaces X_0, X_1, X_2, X_3 have different homotopy types, we prove the following.

Lemma 4.2. *One has*

$$\pi_6(X_n) \cong \mathbf{Z}_6/n\mathbf{Z}_6 \cong \begin{cases} \mathbf{Z}_6 & \textit{for } n = 0 \\ 0 & \textit{for } n = 1 \\ \mathbf{Z}_2 & \textit{for } n = 2 \\ \mathbf{Z}_3 & \textit{for } n = 3\,. \end{cases}$$

Proof. The homotopy sequence of $SU(3) \to X_n \to S^7$ gives

$$\ldots \to \pi_7(S^7) \xrightarrow{\partial} \pi_6(SU(3)) \to \pi_6(X_n) \to \pi_6(S^7) = 0 \to \ldots$$

and if $\hat{\alpha} \in \pi_6(\mathrm{SU}(3))$ denotes the adjoint of $\alpha \in \pi_7(\mathrm{BSU}(3))$, then $\partial(1) = n\hat{\alpha}$. But $\hat{\alpha}$ generates $\pi_6(\mathrm{SU}(3))$ and hence $\pi_6(X_n) = \mathbf{Z}_6/n\mathbf{Z}_6$.

References

[1] R.Bott, A report on the unitary group, Proc. Symp. in Pure Math. III (1961) 1–7.

[2] R.R. Douglas, P.J. Hilton and F. Sigrist, *H*-spaces, Springer Lecture Notes 92 (1969) 65–73.

[3] P. Hilton and J. Roitberg, On principal S^3-bundles over spheres, Ann. Math. 90 (1) (1969) 91–117.

[4] I.M. James and J.H.C. Whitehead, The homotopy theory of sphere bundles over spheres I, Proc. London Math. Soc. 4 (1954) 196–218.

[5] I.M. James and J.H.C. Whitehead, The homotopy theory of sphere bundles over spheres II, Proc. London Math. Soc. 5 (1955) 148–166.

[6] J.-P. Serre, Quelques calculs de groupes d'homotopie, C.R. Paris 236 (1953) 2475–2477.

[7] J. Stasheff, Manifolds of the homotopy type of (non-Lie) groups, Bull. Am. Math. Soc. 75 (1969) 998–1000.

[8] H. Toda, Generalized Whitehead products and homotopy groups of spheres, J. Inst. Polytech. Osaka City Univ. 3 (1952) 43–82.

[9] A. Zabrodsky, Homotopy associativity and finite CW complexes, Memeographed notes, Univ. of Illinois, Chicago Circle.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 1 (1971) pp. 41–77.

AXIOMATIC REPRESENTATION THEORY FOR FINITE GROUPS

J.A. GREEN
Mathematics Institute, University of Warwick, Coventry

Received 25 February 1970

Introduction

The purpose of this paper * is to give a general setting for a part of the representation theory of finite groups, namely the part which includes the theory of *vertices* of modular representations [11, 12] and of R.Brauer's *defect groups* of blocks [3]. In particular we give a general "transfer theorem" (Theorem 2, §4.2) which generalizes the transfer theorem for modules ** of [14] and D.L.Johnson's transfer theorem for cohomology of finite groups [17], as well as Brauer's "first main theorem" for blocks [3, 10B].

For a given finite group G and commutative coefficient ring k, we define "G-functors over k" (§§1.3, 1.8). Such a G-functor $\mathbf{A}$ assigns to each subgroup H of G a k-algebra A_H; to each pair H, K ($H \leqslant K$) of subgroups of G it assigns two k-module homomorphisms

$$T_{H,K} : A_H \to A_K \quad \text{and} \quad R_{K,H} : A_K \to A_H ;$$

and to each subgroup H, and each element g of G, $\mathbf{A}$ also assigns a k-algebra isomorphism $C_{H,g} : A_H \to A_{Hg}$ ($H^g = g^{-1}Hg$). These maps satisfy certain axioms (see §1.3 and §1.8) which embody some familiar landmarks of character theory.

Each G-functor $\mathbf{A}$ defines a "canonical filtration" of ideals of the k-algebras A_H: if $\mathfrak{U}$ is a non-empty set of subgroups of H, $A_{\mathfrak{U},H} = \Sigma \operatorname{Im} T_{U,H}$ (sum over $U \in \mathfrak{U}$) is an ideal of A_H (§2.2), and then the family $(A_{\mathfrak{U},H})$ (for all $\mathfrak{U}$, H) is the canonical filtration on $\mathbf{A}$. In particular, if $\mathfrak{U}$ is a set of subgroups of G, $\mathbf{A}$ is said to be $\mathfrak{U}$-*projective* if $A_{\mathfrak{U},G} = A_G$. Theorem 1 (§3.1) says that if $A_G \cdot A_G = A_G$, then there is a unique "defect base" of $\mathbf{A}$, which is a set $\mathfrak{D} = \mathfrak{D}(\mathbf{A})$ of subgroups of G, closed to taking subgroups and conjugates in G, such that if $\mathfrak{U}$ is another such "j_G-closed" set,

* The author is glad to thank the Master and Fellows of Corpus Christi College, Cambridge, for their kind support and hospitality while this paper was being written.

** Theorem 2 does not give the transfer theorem 1 of [14] directly, see §5.2.

then $\mathbf{A}$ is $\mathfrak{U}$-projective if and only if $\mathfrak{U} \supseteq \mathfrak{T}$.* The classical defect base is Brauer's set $\mathfrak{E}$ of all elementary subgroups of G, which is the defect base of the "character ring functor" on G (§5.1). Other examples correspond to vertices (§5.5a) and defect groups of blocks (§5.5b). Most of section 3 is concerned with techniques for calculating defect bases.

The transfer theorem (Theorem 2) is proved in §4.2, and has as corollary (Proposition 4.34) a generalized version of the module correspondence in [14].

In section 5 are described examples of G-functors drawn from the following theories: §5.1 Characters, §5.2 Grothendieck rings, §5.3 Cohomology, §5.4 K-theory, §5.5 G-algebras (in the sense of [15]). It will be evident that the ideas in this paper have existed in special forms for a long time, e.g. in Tate's rules for calculating cohomology of finite groups ([5], Chapter XII), the proofs of Roquette [22] and Brauer and Tate [4] of Brauer's "characterization of characters" theorem, and Atiyah's functorial description of character theory in [1]. T.Y.Lam [18] has given an axiomatic system which has much in common with ours.

Notation. Maps and functors are usually on the right, with the corresponding convention for writing composites. If G, H are groups (or modules), $H \leqslant G$ means that H is a subgroup (or submodule) of G. If G, H are groups and $H \leqslant G$, $(G : H)$ denotes the index of H in G; an *H-transversal of G* is a set, containing the unit element of G, of representatives of the cosets Hg $(g \in G)$; if $H, K \leqslant G$, an *(H,K)-transversal of G* is a set, containing the unit element of G, of representatives of the double cosets HgK $(g \in G)$. If $x, g \in G$, we write $x^g = g^{-1}xg$; if also $H \leqslant G$, we write $H^g = g^{-1}Hg$.

* A.Dress [8] has anticipated this notion for kG-modules, see §5.2.

§ 1. The subgroup category. *G*-functors

1.1. Let G be a group, with unit element 1. The *subgroup category of* G is defined to be the category $\mathcal{S}(G)$ whose objects are all the subgroups $H, K, L, \ldots$ of G, and for which the set of morphisms from H to K is the set of all triples (H, g, K) such that g is an element of G and $g^{-1}Hg = H^g \leqslant K$. The product of morphisms $(H, g, K)(K, g', L)$ is defined to be (H, gg', L). The identity morphism on H is $(H, 1, H)$. A triple $(H, 1, K)$ is a morphism if and only if $H \leqslant K \leqslant G$; we call such a morphism an *inclusion* *. A morphism (H, g, K) is an isomorphism of $\mathcal{S}(G)$ if and only if $K = H^g$. The following proposition is easily verified:

Proposition 1.11. *A morphism* (H, g, K) *of* $\mathcal{S}(G)$ *can be uniquely factorised as product of an isomorphism by an inclusion, and also as product of an inclusion by an isomorphism, viz.*

$$(H, g, K) = (H, g, H^g)(H^g, 1, K) = (H, 1, K^{g^{-1}})(K^{g^{-1}}, g, K)\,.$$

We shall refer to the morphisms of $\mathcal{S}(G)$ as *G-morphisms*.

1.2. Let k be a commutative ring with identity element. A *k-algebra* is a pair (P, m), where P is a (left, unital) k-module and $m : P \times P \to P$ is a k-bilinear multiplication on P. We write $(\alpha, \beta)m = \alpha\cdot\beta \;\; (\alpha, \beta \in P)$.

If (P, m), (P', m') are k-algebras, then a *k-map* from (P, m) into (P', m') is, by definition, a k-module homomorphism $\theta : P \to P'$; θ is not assumed to be multiplicative. Let $\mathcal{A}_k$ be the category whose objects are all k-algebras, and whose morphisms are all k-maps in the sense just described.

$\mathcal{A}_k$ has a subcategory $\mathcal{M}_k$, whose objects are all k-algebras (P, m) "with zero multiplication", i.e. for which $\alpha\cdot\beta = 0$, for all α, β in P. The morphisms of $\mathcal{M}_k$ are, as before, all k-maps between such objects. Evidently $\mathcal{M}_k$ can be identified with the usual category of k-modules and k-maps.

Notation. If (P, m) is a k-algebra, we make the usual abbreviation and refer to (P, m) as P. The identity map on P is written $\mathrm{id}(P)$.

1.3. *G-functors*

From now on, we assume that G is finite.

Definition. A *G-functor* $\mathbf{A}$ into $\mathcal{A}_k$ (or a *G-functor over k*) written $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_k$, is a triple $\mathbf{A} = (A, T, R)$, where A is a function which assigns to each subgroup H of G a k-algebra A_H, and T, R are functions which assign to each G-morphism $\pi = (H, g, K)$ the respective k-maps

* Each morphism $\pi = (H, g, K)$ defines a monomorphism of groups $(\pi) : h \mapsto h^g (h \in H)$. However, we do *not* identify π with $(\pi) : H \to K$.

$$T(\pi): A_H \to A_K, \quad R(\pi): A_K \to A_H \ ,$$

in such a way that the following axioms hold:

G1 $(A, T): \mathcal{S}(G) \to \mathcal{A}_k$ and $(A, R): \mathcal{S}(G) \to \mathcal{A}_k$ are ordinary functors, covariant and contravariant respectively, i.e. if $\pi = (H, g, K)$, $\pi' = (K, g', L)$ are G-morphisms, then $T(\pi\pi') = T(\pi)\,T(\pi')$ and $R(\pi\pi') = R(\pi')\,R(\pi)$; also $T(H, 1, H) = R(H, 1, H) = \mathrm{id}(A_H)$.

G2 $T(H, g, H^g) = R(H^g, g^{-1}, H)$ for all $H \leqslant G$, $g \in G$.

G3 $T(H, h, H) = R(H, h^{-1}, H) = \mathrm{id}(A_H)$ for all $H \leqslant G$, $h \in H$.

G4 (Mackey axiom) For all subgroups H, K, L of G such that $H \leqslant L$, $K \leqslant L$

$$T(H, 1, L)\,R(K, 1, L) = \sum_{g \in \Gamma} R(H^g \cap K, g^{-1}, H)\,T(H^g \cap K, 1, K) \,,$$

where Γ is an (H, K)-transversal of L.

G5 (Frobenius axiom) For all G-morphisms $\pi = (H, g, K)$ and for all $\alpha \in A_H$, $\beta \in A_K$,

$$\alpha T(\pi) \cdot \beta = (\alpha \cdot \beta R(\pi))\,T(\pi) \ \text{ and } \ \beta \cdot \alpha T(\pi) = (\beta R(\pi) \cdot \alpha)\,T(\pi) \,.$$

Remarks. (1) Axiom G4 does not depend on the choice of Γ. For let $t(g) = R(H^g \cap K, g^{-1}, H)\,T(H^g \cap K, 1, K)$. Then it follows from G1, G2, G3, that $t(hgk) = t(g)$, for all $h \in H$, $g \in L$, $k \in K$. Therefore the sum in G4 is not changed if Γ is changed to another (H, K)-transversal of L.

(2) If all the A_H are objects of $\mathcal{M}_k$, we may regard $\mathbf{A}$ as a *G-functor into* $\mathcal{M}_k$. Axiom G5 is redundant in this case.

1.4. We define next two special types of G-functors.

Definition. A G-functor $\mathbf{A}: \mathcal{S}(G) \to \mathcal{A}_k$ is called *cohomological* if it satisfies the Axiom C:

C For all subgroups H, K of G such that $H \leqslant K$

$$R(H, 1, K)\,T(K, 1, H) = (K : H)\,\mathrm{id}(A_K) \,.$$

Definition. A G-functor $\mathbf{A}: \mathcal{S}(G) \to \mathcal{A}_k$ is called *multiplicative* if it satisfies the Axiom M:

M $R(\pi): A_K \to A_H$ is a multiplicative k-map (i.e. is a homomorphism of k-algebras in the usual sense), for all G-morphisms $\pi = (H, g, K)$.

Remark All "natural" examples of G-functors are multiplicative. For arbitrary G-functors, the Frobenius axiom G5 is a partial substitute for multiplicativity.

Proposition 1.41. *Let* $\mathbf{A}: \mathcal{S}(G) \to \mathcal{A}_k$ *be a G-functor, and let* $\pi = (H, g, H^g)$ *be any isomorphism of* $\mathcal{S}(G)$. *Then* $R(\pi): A_{H^g} \to A_H$ *is a multiplicative isomorphism of k-algebras.*

Proof. By G1 and G2, $T(\pi) = R(\pi^{-1}) = R(\pi)^{-1}$ is a bijective k-map, and so therefore is $R(\pi)$. Let γ, β be any elements of A_{H^g}, and put $\alpha = \gamma R(\pi)$. Then by G5, $\gamma \cdot \beta = (\gamma R(\pi) \cdot \beta R(\pi)) R(\pi)^{-1}$, i.e. $(\gamma \cdot \beta) R(\pi) = \gamma R(\pi) \cdot \beta R(\pi)$. This proves that $R(\pi)$ is a multiplicative isomorphism, as required.

1.5. *The categories* $\mathcal{A}_k(G), \mathcal{M}_k(G)$

In this paragraph, k and G are fixed, and so is the following notation: $\mathbf{A} = (A, T, R)$, $\mathbf{A}' = (A', T', R')$, $\mathbf{A}'' = (A'', T'', R'')$ denote G-functors into $\mathcal{A}_k$; H, K denote subgroups of G, and g an element of G.

A *morphism of G-functors* $\theta: \mathbf{A} \to \mathbf{A}'$ is defined to be a family $(\theta_H)_{H \leqslant G}$ of *multiplicative* k-maps $\theta_H: A_H \to A'_H$ such that

$$(1.51) \qquad \theta_H T'(\pi) = T(\pi)\theta_K \quad \text{and} \quad R(\pi)\theta_H = \theta_K R'(\pi),$$

for all G-morphisms $\pi = (H, g, K)$; i.e. θ must be a natural map, both of (A, T) into (A', T'), and of (A, R) into (A', R'). We denote by $\mathcal{A}_k(G)$ the category whose objects are all G-functors $\mathbf{A}: \mathcal{S}(G) \to \mathcal{A}_k$, and with morphisms as just defined. $\mathcal{M}_k(G)$ denotes the subcategory of $\mathcal{A}_k(G)$ whose objects are all G-functors $\mathbf{A}: S(G) \to \mathcal{M}_k$, and with morphisms defined in the same way. Of course, for a morphism $\theta = (\theta_H)_{H \leqslant G}$ in $\mathcal{M}_k(G)$, the condition that the θ_H be multiplicative is redundant.

If $\mathbf{A}$, $\mathbf{A}'$ are objects of $\mathcal{A}_k(G)$, we say that $\mathbf{A}'$ is a *subfunctor [ideal] of* $\mathbf{A}$, and write $\mathbf{A}' \leqslant \mathbf{A}$ $[\mathbf{A}' \trianglelefteq \mathbf{A}]$, if $\mathbf{A}'$ satisfies conditions (1.52a, b, c) below:

(1.52a) For each subgroup H of G, A'_H is a k-subalgebra [k-ideal] of A_H.

(1.52b) For each G-morphism $\pi = (H, g, K)$, $A'_H T(\pi) \leqslant A'_K$ and $A'_K R(\pi) \leqslant A'_H$.

(1.52c) For each G-morphism $\pi = (H, g, K)$, $T'(\pi), R'(\pi)$ are the restrictions of $T(\pi), R(\pi)$ to $A'_H \to A'_K$, $A'_K \to A'_H$ respectively.

Conversely, any family $A' = (A'_H)_{H \leqslant G}$ of k-algebras, which satisfies (1.52a, b), defines a unique subfunctor [ideal] $\mathbf{A}'$ of $\mathbf{A}$, by taking (1.52c) as definition of the functions T', R'.

If $\mathbf{A}'$ is an ideal of $\mathbf{A}$, we define the *quotient functor* $\mathbf{A}'' = \mathbf{A}/\mathbf{A}'$ by saying that $A''_H = A_H/A'_H$ for all subgroups H of G, and that for each G-morphism $\pi = (H, g, K)$, $T''(\pi)$ and $R''(\pi)$ are given by

$$T''(\pi): \alpha + A'_H \mapsto \alpha T(\pi) + A'_K, \quad \alpha \in A_H,$$

and

$$R''(\pi) : \beta + A'_K \mapsto \beta R(\pi) + A'_H \,, \quad \beta \in A_K \,,$$

respectively.

Every morphism of G-functors $\theta : \mathbf{A} \to \mathbf{A}'$ has *kernel* $\operatorname{Ker}\theta$, an ideal of $\mathbf{A}$, and *image* $\operatorname{Im}\theta$, a subfunctor of $\mathbf{A}'$; the definitions of these, and of the "canonical" isomorphism of G-functors

$$\bar{\theta} : \mathbf{A}/\operatorname{Ker}\theta \to \operatorname{Im}\theta$$

are straightforward. The category $\mathcal{A}_k(G)$ has zero objects, namely all "zero G-functors" $\mathbf{0} : \mathcal{S}(G) \to \mathcal{A}_k$, i.e. G-functors for which $A_H = 0_H$, a zero k-algebra, for all H. Any zero G-functor belongs to $\mathcal{M}_k$. Exact sequences and short exact sequences can be defined in $\mathcal{A}_k(G)$ in the expected way.

If $\mathbf{A}$ is an object of $\mathcal{M}_k(G)$, every subfunctor of $\mathbf{A}$ is also an ideal of $\mathbf{A}$. Therefore every morphism in $\mathcal{M}_k(G)$ has both kernel and cokernel; $\mathcal{M}_k(G)$ is an Abelian category. In §5.6 we describe certain "relatively free" objects in $\mathcal{M}_k(G)$.

1.6. *Change of coefficient ring*

Let r be a commutative ring with identity element 1_r, and $\phi : k \to r$ a ring homomorphism with $1_k\phi = 1_r$. Then ϕ induces a covariant, additive functor $\Phi : \mathcal{M}_k \to \mathcal{M}_r$ which takes each k-module P to the r-module $P\Phi = r \underset{k}{\otimes} P$ (r is regarded as left r-module by left multiplication, and as right k-module by means of ϕ. See e.g. [5] p. 29). We extend Φ to a covariant, additive functor $\Phi : \mathcal{A}_k \to \mathcal{A}_r$ as follows. If $P = (P, m)$ is a k-algebra with multiplication $m : (\alpha, \beta) \to \alpha \cdot \beta$, then $(P, m)\Phi$ is defined to be $(P\Phi, m\Phi)$, where $m\Phi$ is the r-bilinear multiplication on $P\Phi$ obtained from the rule

$$(a \otimes \alpha) \cdot (b \otimes \beta) = ab \otimes (\alpha \cdot \beta) \qquad (a, b \in r,\ \alpha, \beta \in P)\,.$$

The action of Φ on k-maps between (P, m) and (P', m') is defined to coincide with the action of Φ already defined on these k-maps, regarded as morphisms in $\mathcal{M}_k$.

If $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_k$ is a G-functor, then $\mathbf{A}\Phi : \mathcal{S}(G) \to \mathcal{A}_r$ is also a G-functor. We write $\mathbf{A}\Phi = r \underset{k}{\otimes} \mathbf{A}$ in contexts where it is not necessary to mention ϕ, Φ explicitly. It is also clear that the correspondence $\mathbf{A} \to \mathbf{A}\Phi = r \underset{k}{\otimes} \mathbf{A}$ between objects of $\mathcal{A}_k(G)$ and $\mathcal{A}_r(G)$, can be extended to give a covariant functor from $\mathcal{A}_k(G)$ to $\mathcal{A}_r(G)$: this functor takes a morphism $(\theta_H)_{H \leqslant G}$ in $\mathcal{A}_k(G)$ to the morphism $(\theta_H\Phi)_{H \leqslant G}$ in $\mathcal{A}_r(G)$.

1.7. *Change of group*

If G_1, G are finite groups and $\gamma : G_1 \to G$ a group homomorphism, then γ induces a covariant functor $\mathcal{S}(\gamma) : \mathcal{S}(G_1) \to \mathcal{S}(G)$, which takes each subgroup H_1 of G_1 to the subgroup $H_1\gamma$ of G_1, and each G_1-morphism $\pi_1 = (H_1, g_1, K_1)$ to the G-morphism

$\pi_1 \mathcal{S}(\gamma) = (H_1\gamma, g_1\gamma, K_1\gamma)$. If $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_k$ is a G-functor into $\mathcal{A}_k$ and *if γ is a monomorphism,* then $\mathcal{S}(\gamma)\mathbf{A} : \mathcal{S}(G_1) \to \mathcal{A}_k$ is a G_1-functor into $\mathcal{A}_k$. The condition that γ be a monomorphism, is to ensure that $(H_1{}^{g_1} \cap K_1)\gamma = (H_1\gamma)^{g_1\gamma} \cap (K_1\gamma)$ for any subgroups H_1, K_1 of G_1 and element g_1 of G_1; this is necessary to make Axiom G4 hold for $\mathcal{S}(\gamma)\mathbf{A}$.

If γ is a monomorphism, the correspondence $\mathbf{A} \to \mathcal{S}(\gamma)\mathbf{A}$ can be extended to a covariant functor of $\mathcal{A}_k(G)$ into $\mathcal{A}_k(G_1)$; this functor takes a morphism $(\theta_H)_{H \leqslant G}$ in $\mathcal{A}_k(G)$ to the morphism $(\theta_{H_1\gamma})_{H_1 \leqslant G_1}$ in $\mathcal{A}_k(G)$.

In the special case that G_1 is a subgroup of G and γ is the inclusion monomorphism, we write $\mathcal{S}(\gamma)\mathbf{A} = {}_{G_1}\mathbf{A}$, and call this G_1-functor the *restriction of* $\mathbf{A}$ *to* G_1.

1.8. *The maps $T_{H,K}$, $R_{K,H}$, $C_{H,g}$*

Let $\mathbf{A} = (A, T, R)$ be a G-functor into $\mathcal{A}_k$. It is possible to express all the maps $T(\pi)$, $R(\pi)$, for all G-morphisms π, in terms of the following special maps.

Definition 1.81. For all subgroups H, K of G such that $H \leqslant K$, and for all elements g in G, define

(a) $\quad T_{H,K} = T(H, 1, K) : A_H \to A_K$.

(b) $\quad R_{K,H} = R(H, 1, K) : A_K \to A_H$.

(c) $\quad C_{H,g} = T(H, g, H^g) = R(H^g, g^{-1}, H) : A_H \to A_{H^g}$.

By Proposition 1.11 and Axioms G1, G2, we have

Proposition 1.82. *If $\pi = (H, g, K)$ is a G-morphism, then*

(a) $\quad T(\pi) = C_{H,g} T_{H^g,K}$.

(b) $\quad R(\pi) = R_{K,H^g} C_{H^g,g^{-1}}$.

In the next proposition, we express the Axioms G1–G5 in terms of the maps $T_{H,K}$, $R_{K,H}$, $C_{H,g}$; this provides a new form for the definition of a G-functor, which, although rather longer than the one we gave in §1.3, is often easier to use.

Proposition 1.83. *If $\mathbf{A} = (A, T, R)$ is a G-functor into $\mathcal{A}_k$, and if $T_{H,K}$, $R_{K,H}$, $C_{H,g}$ are defined by Definition 1.81, then the following equations hold for all subgroups D, H, K, L of G and all elements g, g′ of G.*

(a) $\quad T_{H,H} = \mathrm{id}(A_H)$, *and* $T_{H,K} T_{K,L} = T_{H,L}$ *if* $H \leqslant K \leqslant L$.

(b) $\quad R_{H,H} = \mathrm{id}(A_H)$, *and* $R_{K,H} R_{H,D} = R_{K,D}$ *if* $K \geqslant H \geqslant D$.

(c) $\quad C_{H,g} C_{H^g,g'} = C_{H,gg'}$.

(d) $\quad C_{H,h} = \mathrm{id}(A_H)$ *if* $h \in H$.

(e) $T_{H,K}C_{K,g} = C_{H,g}T_{Hg,Kg}$.

(f) $R_{K,H}C_{H,g} = C_{K,g}R_{Kg,Hg}$.

(g) $T_{H,L}R_{L,K} = \sum_{g \in \Gamma} C_{H,g}R_{Hg,Hg\cap K}\,T_{Hg\cap K,K}$,
if $H \leqslant L$, $K \leqslant L$ *and* Γ *is an* (H, K)*-transversal of* L.

(h) $C_{H,g} : A_H \to A_{Hg}$ *is a multiplicative isomorphism of k-algebras.*

(i) $\alpha T_{H,K} \cdot \beta = (\alpha \cdot \beta R_{K,H})T_{H,K}$ and $\beta \cdot \alpha T_{H,K} = (\beta R_{K,H} \cdot \alpha)\,T_{H,K}$ *if* $H \leqslant K$, $\alpha \in A_H$ *and* $\beta \in A_K$.

Conversely, suppose that k-algebras A_H *and k-maps* $T_{H,K} : A_H \to A_K$, $R_{K,H} : A_K \to A_H$, $C_{H,g} : A_H \to A_{Hg}$ *are given, for all subgroups* H, K *of* G *such that* $H \leqslant K$, *and all elements* g *of* G, *in such a way that* 1.83(a)–(i) *hold. For each G-morphism* $\pi = (H, g, K)$ *define* $T(\pi)$, $R(\pi)$ *by* 1.82(a), (b) *respectively. Then* $\mathbf{A} = (A, T, R)$ *is a G-functor into* $\mathscr{A}_k$.

Proof. (a), (b), (c) follow from G1, (d) from G3, (e) and (f) by applying T and R respectively to suitable cases of Proposition 1.11, (g) follows from G4 and G1, (h) from Proposition 1.41, and (i) from G5, taking $\pi = (H, 1, K)$.

For the converse, the proof consists in verifying that axioms G1–G5 hold, when T, R are defined by 1.82, it being given that conditions (a)–(i) hold. We leave this to the reader.

Proposition 1.84. *Let* $\mathbf{A} = (A, T, R)$ *be as before, and let* H, K, L *be subgroups of* G *such that* $H \leqslant L$, $K \leqslant L$. *Then for all* $\alpha \in A_H$, $\beta \in A_K$,

$$\alpha T_{H,L} \cdot \beta T_{K,L} = \sum_{g \in \Gamma} (\alpha C_{H,g} R_{Hg,Hg\cap K} \cdot \beta R_{K,Hg\cap K})\,T_{Hg\cap K,L}\,,$$

where Γ *is an* (H, K)*-transversal of* L.

Proof. $\alpha T_{H,L} \cdot \beta T_{K,L} = (\alpha T_{H,L} R_{L,K} \cdot \beta)\,T_{K,L}$, (by 1.83(i))

$$= \sum_{g \in \Gamma} (\alpha C_{H,g} R_{Hg,Hg\cap K}\, T_{Hg\cap K,K} \cdot \beta)\, T_{K,L}\,, \quad \text{(by 1.83(g))}$$

$$= \sum_{g \in \Gamma} (\alpha C_{H,g} R_{Hg,Hg\cap K} \cdot \beta R_{K,Hg\cap K})\, T_{Hg\cap K,K} T_{K,L}\,, \quad \text{(by 1.83(i))}\,,$$

and the proof is completed by observing that $T_{Hg\cap K,K}T_{K,L} = T_{Hg\cap K,L}$, by 1.83(a).

Remarks. 1) The condition that $\mathbf{A} = (A, T, R)$ be cohomological can be written: $R_{H,K}T_{K,H} = (K : H)\,\mathrm{id}(A_K)$, for all subgroups H, K of G such that $H \leqslant K$.

2) The condition that $\mathbf{A} = (A, T, R)$ be multiplicative is that $R_{K,H} : A_K \to A_H$ be

a multiplicative k-map, for all subgroups H, K of G such that $H \leqslant K$. For we know that the $C_{H,g}$ are always multiplicative (Proposition 1.41 or 1.83(h)), for any G-functor **A**. By 1.82(b), the condition that the $R_{K,H}$ be multiplicative then ensures that all the $R(\pi)$ are multiplicative.

3) Let $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_k$ be a G-functor, and H a fixed subgroup of G. For any α in A_H and g in G, write $\alpha^g = \alpha C_{H,g}$. If g lies in $N(H)$, the normalizer of H in G, then α^g is in A_H. Using 1.83(c), (h) we see that this action of $N(H)$ on A_H makes A_H into a $kN(H)$-module, and if A_H has an identity element, A_H is even an $N(H)$-*algebra* over k in the sense of [15] p. 138. Moreover H acts trivially on A_H by 1.83(d), so the action of $N(H)$ can be "lifted" to an action of $N(H)/H$ on A_H.

§ 2. The canonical filtration on a G-functor

2.1. We make, in this paragraph, some definitions concerning sets of subgroups of a group G. If U, V are subgroups of H, where H is a given subgroup of G, we write $U \leqslant_H V$ to mean that there exists some h in H such that $U^h \leqslant V$. If $\mathfrak{U}$, $\mathfrak{V}$ are *sets* of subgroups of H, we write $\mathfrak{U} \leqslant_H \mathfrak{V}$ to mean that for each U in $\mathfrak{U}$ there exists some V in $\mathfrak{V}$ such that $U \leqslant_H V$; and $\mathfrak{U} =_H \mathfrak{V}$ to mean that $\mathfrak{U} \leqslant_H \mathfrak{V}$ and $\mathfrak{V} \leqslant_H \mathfrak{U}$. It is useful to express these relations $\leqslant_H$, $=_H$ in terms of a "closure operation" j_H as follows.

We make the convention that, from now on, "set of subgroups" shall always mean "non-empty set of subgroups".

Definition. If $\mathfrak{U}$ is a set of subgroups of H, define

$$j_H\mathfrak{U} = \{V \leqslant H \mid \exists U \in \mathfrak{U} \text{ such that } V \leqslant_H U\}\,,$$

and call $j_H\mathfrak{U}$ the *j_H-closure of* $\mathfrak{U}$. The set $\mathfrak{U}$ is said to be *j_H-closed* if $j_H\mathfrak{U} = \mathfrak{U}$; thus $\mathfrak{U}$ is j_H-closed if and only if $\mathfrak{U}$ contains, with any subgroup $U \in \mathfrak{U}$, all the subgroups of U and all the conjugates of U in H.

Proposition 2.11. *Let $\mathfrak{U}$, $\mathfrak{V}$ be sets of subgroups of H. Then*

(a) $\mathfrak{U} \subseteq j_H\mathfrak{U}$.

(b) $j_H(j_H\mathfrak{U}) = j_H\mathfrak{U}$.

(c) $\mathfrak{U} \leqslant_H \mathfrak{V}$ if and only if $j_H\mathfrak{U} \subseteq j_H\mathfrak{V}$; hence $\mathfrak{U} =_H \mathfrak{V}$ if and only if $j_H\mathfrak{U} = j_H\mathfrak{V}$.

(d) *Every j_H-closed $\mathfrak{U}$ contains the unit subgroup $\{1\}$ of H.*

(e) *The intersection and union of any non-empty set of j_H-closed sets of subgroups of H are both j_H-closed sets of subgroups of H.*

We omit the proof of Proposition 2.11, which is elementary.

Definition. Let H, K, L be subgroups of G such that $H \leqslant L$, $K \leqslant L$, and let $\mathfrak{U}$, $\mathfrak{V}$ be sets of subgroups of H, K respectively. Define

$$\mathfrak{U} : L : \mathfrak{V} = \{U^g \cap V \mid U \in \mathfrak{U},\ V \in \mathfrak{V},\ g \in L\}\,.\ *$$

Proposition 2.12. *With the notation of the last definition,*

$$j_L(\mathfrak{U} : L : \mathfrak{V}) = j_L\mathfrak{U} \cap j_L\mathfrak{V}\,.$$

* If $\mathfrak{U}$ or $\mathfrak{V}$ consists of a single subgroup, we omit the brackets $\{\ \}$ from $\mathfrak{U} : L : \mathfrak{V}$, e.g. we write $U : L : \mathfrak{V}$ instead of $\{U\} : L : \mathfrak{V}$.

Proof. *If* $X \in j_L(\mathfrak{U} : L : \mathfrak{B})$, there exist $a, b \in L$ and $U \in \mathfrak{U}$, $V \in \mathfrak{B}$ such that $X^a \leqslant U^b \cap V$. Hence $X \leqslant_L U$, $X \leqslant_L V$, and therefore $X \in j_L \mathfrak{U} \cap j_L \mathfrak{B}$. Conversely, if $Y \in j_L \mathfrak{U} \cap j_L \mathfrak{B}$, there exist $a, b \in L$ and $U \in \mathfrak{U}$, $V \in \mathfrak{B}$ such that $Y^a \leqslant U$, $Y^b \leqslant V$. Hence $Y^b \leqslant U^{a^{-1}b} \cap V \in \mathfrak{U} : L : \mathfrak{B}$ and therefore $Y \in j_L(\mathfrak{U} : L : \mathfrak{B})$. This completes the proof.

2.2. *The canonical filtration*

Definition. Let $\mathbf{A} = (A, T, R)$ be a G-functor into $\mathscr{A}_k$. For each pair $(\mathfrak{U}, H)$, where H is a subgroup of G and $\mathfrak{U}$ is a set of subgroups of H, define

$$A_{\mathfrak{U},H} = \sum_{U \in \mathfrak{U}} A_U T_{U,H} \,.$$

The family $(A_{\mathfrak{U},H})$, indexed by the set of pairs $(\mathfrak{U}, H)$ described above, is called the *canonical filtration on* $\mathbf{A}$.

Notation. If $\mathfrak{U}$ consists of a single subgroup U, we write $A_{\mathfrak{U},H} = A_{U,H}$, rather than $A_{\{U\},H}$.

Proposition 2.21. *Let* $\mathbf{A} : \mathscr{S}(G) \to \mathscr{A}_k$ *be a G-functor, and* $(A_{\mathfrak{U},H})$ *its canonical filtration. Then each* $A_{\mathfrak{U},H}$ *is an ideal of* A_H, *and the following statements are true for all subgroups D, H, K of G, and all sets* $\mathfrak{U}$, $\mathfrak{B}$ *of subgroups of H:*

(a) $A_{H,H} = A_H$.

(b) $\mathfrak{U} \leqslant_H \mathfrak{B}$ *implies* $A_{\mathfrak{U},H} \leqslant A_{\mathfrak{B},H}$.

(c) $A_{\mathfrak{U},H} \cdot A_{\mathfrak{B},H} \leqslant A_{\mathfrak{U} : H : \mathfrak{B} : H}$.

(d) $A_{\mathfrak{U},H} T_{H,K} = A_{\mathfrak{U},K}$ *if* $H \leqslant K$.

(e) $A_{\mathfrak{U},H} R_{H,D} \leqslant A_{\mathfrak{U} : H : \{D\}, D}$ *if* $H \geqslant D$.

(f) $A_{\mathfrak{U},H} C_{H,g} = A_{\mathfrak{U}^g, H^g}$ *for all g in G, and* $\mathfrak{U}^g = \{U^g \mid U \in \mathfrak{U}\}$.

Proof. Each set $A_U T_{U,H}$ $(U \leqslant H)$ is an ideal of A_H by Axiom G5. Therefore $A_{\mathfrak{U},H}$ is an ideal of A_H. (a) holds because $T_{H,H} = \mathrm{id}(A_H)$, by 1.83(a). From 1.83(a) we also deduce that

(g) $A_{U,H} T_{H,K} = A_{U,K}$,

whenever $U \leqslant H \leqslant K$, and this implies (d). By 1.83(e) we get

(h) $A_{U,H} C_{H,g} = A_{U^g, H^g}$,

for any $U \leqslant H$ and g in G, and this implies (f). Now (g), (h) together give that $U \leqslant_H V$ implies $A_{U,H} \leqslant A_{V,H}$, for any subgroups U, V of H, and (b) follows from this. (c) is a direct consequence of Proposition 1.84. Finally (e) follows from 1.83(g).

§ 3. Defect bases

3.1. Let $\mathbf{A}: \mathcal{S}(G) \to \mathcal{A}_k$ be a fixed G-functor, and $(A_{\mathfrak{U},H})$ the canonical filtration on $\mathbf{A}$. In this paragraph we are concerned only with the part of the family $(A_{\mathfrak{U},H})$ for which $H = G$; thus $\mathfrak{U}, \mathfrak{V}$ will refer to sets of subgroups of G, and we write $j = j_G$.

Definition. If $\mathfrak{U}$ is a set of subgroups of G, we say that A_G is $\mathfrak{U}$-*projective* if $A_{\mathfrak{U},G} = A_G$; we also say that $\mathbf{A}$ itself is $\mathfrak{U}$-projective in this case.

Lemma 3.11. (a) *A_G is $\{G\}$-projective.*

(b) *A_G is $\mathfrak{U}$-projective if and only if it is $j\mathfrak{U}$-projective.*

(c) *If A_G is $\mathfrak{U}$-projective, and if $\mathfrak{U} \leqslant_G \mathfrak{V}$, then A_G is $\mathfrak{V}$-projective.*

(d) *If $A_G \cdot A_G = A_G$ and if A_G is both $\mathfrak{U}$-projective and $\mathfrak{V}$-projective, then A_G is $(j\mathfrak{U} \cap j\mathfrak{V})$-projective.*

Proof. (a) follows from Proposition 2.21(a). From 2.21(b) we deduce that if $\mathfrak{U} \leqslant_G \mathfrak{V}$ then $A_{\mathfrak{U},G} = A_G$ implies $A_{\mathfrak{V},G} = A_G$, and this proves (c). Since $\mathfrak{U} =_G j\mathfrak{U}$, it also proves (b). Finally, if the hypotheses of (d) hold, then by 2.21(c)

$$A_G = A_G \cdot A_G = A_{\mathfrak{U},G} \cdot A_{\mathfrak{V},G} \leqslant A_{\mathfrak{U}:G:\mathfrak{V},G},$$

so that A_G is $\mathfrak{U}{:}G{:}\mathfrak{V}$-projective. By (b) and Proposition 2.12, A_G is $(j\mathfrak{U} \cap j\mathfrak{V})$-projective. This proves the lemma.

Definition. Let $\mathfrak{D}$ be a set of subgroups of G. Then we say that $\mathfrak{D}$ is a *defect base* for A_G, or for the G-functor $\mathbf{A}$, if

D1: For all $\mathfrak{U}$, A_G is $\mathfrak{U}$-projective if and only if $\mathfrak{D} \leqslant_G \mathfrak{U}$, and

D2: $\mathfrak{D}$ is j-closed.

Theorem 1. *Let $\mathbf{A}: \mathcal{S}(G) \to \mathcal{A}_k$ be a G-functor such that $A_G \cdot A_G = A_G$. Then A_G has a unique defect base.*

Proof. Let Π be the set of all j-closed sets $\mathfrak{U}$ of subgroups of G such that A_G is $\mathfrak{U}$-projective. Π is not empty, since by 3.11(a), (b) it contains $j\{G\}$, the set of all subgroups of G. By 3.11(d) and 2.11(e), the intersection $\mathfrak{D}$ of all the members of Π is itself a member of Π. If $\mathfrak{D} \leqslant_G \mathfrak{U}$ for some $\mathfrak{U}$, then A_G is $\mathfrak{U}$-projective, by 3.11(c). Conversely, if A_G is $\mathfrak{U}$-projective for some $\mathfrak{U}$, then $j\mathfrak{U} \in \Pi$, so $\mathfrak{D} \subseteq j\mathfrak{U}$, i.e. $\mathfrak{D} \leqslant_G \mathfrak{U}$, by 2.11(c). Therefore $\mathfrak{D}$ satisfies D1, hence is a defect base for A_G. The uniqueness of $\mathfrak{D}$ follows immediately from the definition of defect base.

Remarks. 1) If A_G has identity element 1_G, then the condition $A_G \cdot A_G = A_G$ is satisfied. If $\mathfrak{U}$ is any set of subgroups of G, then A_G is $\mathfrak{U}$-projective if and only if $1_G \in A_{\mathfrak{U},G}$, since $A_{\mathfrak{U},G}$ is an ideal of A_G.

2) Examples can be constructed to show that, for any given j-closed set $\mathfrak{U}$ of sub-

groups of G, there is a G-functor $\mathbf{A}$ with $\mathfrak{U}$ as defect base. However, if $\mathbf{A}$ is cohomological (§1.4), then the defect base of $\mathbf{A}$ is strongly restricted.

Proposition 3.12. *Let* $\mathbf{A}: \mathcal{S}(G) \to \mathcal{A}_k$ *be a cohomological G-functor such that* $A_G \cdot A_G = A_G$, *and let* $\pi(k, G)$ *be the set of rational primes p dividing* $|G|$ *and such that* $p1_k$ *is not a unit in k. Then each member of the defect base* $\mathfrak{T}$ *of* A_G *is a p-subgroup of G, for some p in* $\pi(k, G)$.

Proof. If U is any subgroup of G, then $A_U T_{U,G}$ contains $A_G R_{G,U} T_{U,G} = (G:U)A_G$, by Axiom C. If $\mathfrak{U}$ is any set of subgroups of G, we have

$$A_{\mathfrak{U},G} = \sum_{U \in \mathfrak{U}} A_U T_{U,G} \geqslant \sum_{U \in \mathfrak{U}} (G:U)A_G = h(\mathfrak{U}) \cdot A_G ,$$

where $h(\mathfrak{U}) = \{\underset{U \in \mathfrak{U}}{\text{h.c.f.}}\,(G:U)\}\, 1_k$. Now take for $\mathfrak{U}$ the j-closed set consisting of all p-subgroups of G, for all p in $\pi(k, G)$. $\mathfrak{U}$ contains, for each p in $\pi(k, G)$, a Sylow p-subgroup G_p of G. The h.c.f. of these $(G:G_p)$, and *a fortiori* also the h.c.f. of all the $(G:U)$, $U \in \mathfrak{U}$, is a product of primes q such that $q1_k$ is a unit in k. Therefore $A_{\mathfrak{U},G} \geqslant h(\mathfrak{U}) \cdot A_G = A_G$, i.e. A_G is $\mathfrak{U}$-projective. Hence $\mathfrak{T} \subseteq \mathfrak{U}$, and the proposition is proved.

3.2. The definition and existence of a defect base for a G-functor $\mathbf{A}$ (i.e., for A_G) depend only on a small number of properties of the family of ideals $A_{\mathfrak{U},G}$ of A_G. In the next definition we isolate these properties, so that we shall be able to speak of the defect base of a k-algebra A relative to a "G-family" of ideals of A. This allows us to develop some methods for calculating defect bases.

Definition. Let A be a k-algebra, G a finite group, and $\mathfrak{s}(G)$ the set of all sets of subgroups of G. Suppose that $(A_{\mathfrak{U}})$ is a family of ideals of A, indexed by $\mathfrak{U}$ in $\mathfrak{s}(G)$, which satisfies the following conditions, for all $\mathfrak{U}, \mathfrak{V}$ in $\mathfrak{s}(G)$:

GF1 $\quad A_{\{G\}} = A$.

GF2 $\quad \mathfrak{U} \leqslant_G \mathfrak{V}$ implies $A_{\mathfrak{U}} \subseteq A_{\mathfrak{V}}$.

GF3 $\quad A_{\mathfrak{U}} \cdot A_{\mathfrak{V}} \leqslant A_{\mathfrak{U}:G:\mathfrak{V}}$.

Then $(A_{\mathfrak{U}})$ is called a *G-family on A*.

For example, the family $(A_{\mathfrak{U},G})$ of A_G, where A_G and the $A_{\mathfrak{U},G}$ are defined by a G-functor $\mathbf{A}$, is a G-family on A_G. To prove this, we have only to notice that Conditions GF1, 2, 3 are the special cases $H = G$ of the assertions of Proposition 2.21(a), (b), (c).

If $(A_{\mathfrak{U}})$ is a G-family on the k-algebra A, we say that A is $\mathfrak{U}$-*projective, relative to*

$(A_{\mathfrak{U}})$, if $A_{\mathfrak{U}} = A$; we define a *defect base* $\mathfrak{D}$ *relative to* $(A_{\mathfrak{U}})$ to be a set $\mathfrak{D}$ of subgroups of G which satisfies conditions analogous to D1, D2. The proofs of Lemma 3.11 and Theorem 1 go over, and we have

Proposition 3.21. *Let A be a k-algebra such that $A \cdot A = A$, and let $(A_{\mathfrak{U}})$ be a G-family on A. Then A has a unique defect base $\mathfrak{D} = \mathfrak{D}(A)$ relative to $(A_{\mathfrak{U}})$.*

3.3. In this paragraph and the next, we assume that A *is an associative k-algebra with identity element* 1. We write $\alpha\beta$ instead of $\alpha \cdot \beta$ for the product of elements α, β in A; rad A denotes the Jacobson radical of A, and $(A_{\mathfrak{U}})$ a fixed G-family on A.

Proposition 3.31. *Let S be a subalgebra * of A. Then $(S \cap A_{\mathfrak{U}})$ is a G-family on S, and $\mathfrak{D}(A) \supseteq \mathfrak{D}(S)$.*

Proof. It is trivial that $(S \cap A_{\mathfrak{U}})$ is a G-family on S. The other assertion follows from the fact that, for any $\mathfrak{U} \in \mathfrak{s}(G)$, $A_{\mathfrak{U}} = A$ implies $A_{\mathfrak{U}} \cap S = S$.

Proposition 3.32. *Let I be an ideal of A. Then*

(a) *$(A_{\mathfrak{U}} + I/I)$ is a G-family on A/I, and $\mathfrak{D}(A) \supseteq \mathfrak{D}(A/I)$.*

(b) *If I lies in* rad A, *then* $\mathfrak{D}(A) = \mathfrak{D}(A/I)$.

Proof. (a) It is trivial that $(A_{\mathfrak{U}} + I/I)$ is a G-family on A/I. The other assertion follows from the fact that, for any $\mathfrak{U} \in \mathfrak{s}(G)$, $A_{\mathfrak{U}} = A$ implies $A_{\mathfrak{U}} + I/I = A/I$.

(b) In view of (a), it is enough to prove that $\mathfrak{D}(A) \subseteq \mathfrak{D}(A/I)$ in the present case. Let $\mathfrak{U} = \mathfrak{D}(A/I)$. Since A/I is U-projective, $1 + I \in A_{\mathfrak{U}} + I/I$, i.e. $1 + \alpha \in A_{\mathfrak{U}}$ for some α in I. But then α is in rad A, and so $1 + \alpha$ has an inverse in A. Hence $1 \in A_{\mathfrak{U}}$, and therefore A is $\mathfrak{U}$-projective. Since $\mathfrak{U}$ is j-closed, $\mathfrak{D}(A) \subseteq \mathfrak{U}$, as required. This completes the proof of Proposition 3.32.

Suppose that e is an idempotent in A. The $S = eAe$ is a subalgebra of A; also $eAe \cap I = eIe$, for any ideal I of A. So the G-family on $eAe = S$ given by Proposition 3.31 is $(eA_{\mathfrak{U}}e)$. For any $\mathfrak{U} \in \mathfrak{s}(G)$, the condition for eAe to be $\mathfrak{U}$-projective is that $e \in eA_{\mathfrak{U}}e$, which is equivalent to the condition $e \in A_{\mathfrak{U}}$.

Proposition 3.33. *If $1 = e_1 + ... + e_n$, where $e_1, ..., e_n$ are idempotents in A, then $\mathfrak{D}(A) = \bigcup_{i=1}^{n} \mathfrak{D}(e_iAe_i)$.*

Proof. Write $\mathfrak{D}_i = \mathfrak{D}(e_iAe_i)$, so that by 3.31, $\mathfrak{D}(A) \supseteq \mathfrak{D}_i$, for all i. Therefore $\mathfrak{D}(A) \supseteq \mathfrak{D} = \bigcup_{i=1}^{n} \mathfrak{D}_i$. But we have also $e_i \in A_{\mathfrak{D}_i}$, for all i, by the remark above. Therefore $1 = \sum e_i \in \sum A_{\mathfrak{D}_i} \leqslant A_{\mathfrak{D}}$, so that A is $\mathfrak{D}$-projective. Since $\mathfrak{D}$ is j-closed (Proposition 2.11(e)), $\mathfrak{D}(A) \subseteq \mathfrak{D}$, and this completes the proof of Proposition 3.33.

* We do not assume that $1 \in S$.

Corollary 3.34. *If $A = A_1 \oplus \dots \oplus A_n$ (direct sum of ideals A_i of A), then $\mathfrak{X}(A) = \bigcup_{i=1}^{n} \mathfrak{X}(A_i)$.*

Proof. Take e_i to be the identity element of A_i, $i = 1, \dots, n$, and use Proposition 3.33. Since $A_i = e_i A e_i$ for all i, this proves the corollary.

If e, f are idempotents of A, we shall say that e, f *are associated in A* if the right A-modules eA, fA are isomorphic. Jacobson ([16] p. 51) shows that e, f are associated in A if and only if there exist elements x, y in A such that

(1) $\qquad exf = x\,,\quad fye = y\,,\quad xy = e\,,\quad yx = f\,.$

It follows that any ideal $A_{\mathfrak{U}}$ of A which contains e, also contains f, and conversely. This proves

Proposition 3.35. *If e, f are idempotents which are associated in A, then $\mathfrak{X}(eAe) = \mathfrak{X}(fAf)$.*

We shall say that an idempotent e of A is *completely primitive in A* if eAe is completely primary in the sense of Jacobson [loc. cit. p. 56], i.e. if $eAe/\mathrm{rad}(eAe)$ is a division algebra.

In the next proposition and definition, we shall assume that the G-family $(A_{\mathfrak{U}})$ satisfies, for all $\mathfrak{U}$ in $\mathfrak{s}(G)$,

GF4 $\qquad A_{\mathfrak{U}} = \sum_{U \in \mathfrak{U}} A_{\{U\}}\,.$

Note that GF4 holds for the G-family $(A_{\mathfrak{U},G})$ defined on A_G by a G-functor $\mathbf{A}$.

Proposition 3.36. *If e is a completely primitive idempotent of A, then $\mathfrak{X}(eAe) = j\{D\}$, for some subgroup D of G, which is determined uniquely up to conjugacy in C.*

Definition. With the notation of Proposition 3.36, we say that D is a *defect group* of e, or of eAe.

Proof of Proposition 3.36. Let $\mathfrak{X} = \mathfrak{X}(eAe)$. Then $e \in A_{\mathfrak{X}} = \sum_{D \in \mathfrak{X}} A_D$.* and so $e \in \sum_{D \in \mathfrak{X}} eA_De$. If all the eA_De were proper ideals of eAe, they would all lie in $\mathrm{rad}(eAe)$ (see Jacobson, loc. cit., Prop. 10.1), giving the contradiction $e \in \mathrm{rad}(eAe)$. So there is some D in $\mathfrak{X}$ for which $e \in A_D$. Therefore eAe is $\{D\}$-projective, hence $\mathfrak{X} \subseteq j\{D\}$. But $D \in \mathfrak{X}$ and $\mathfrak{X}$ is j-closed, so $j\{D\} \subseteq \mathfrak{X}$. This completes the proof, for it is clear that D is determined uniquely up to conjugacy in G by the equation $\mathfrak{X} = j\{D\}$.

* A_D stands for $A_{\{D\}}$.

In order to apply these results G-functors, we need

Lemma 3.37. *Let* $\mathbf{A}:\mathcal{S}(G)\to\mathcal{A}_k$ *be a multiplicative G-functor such that each algebra* A_H *is associative. Let* $e=e_G$ *be an idempotent of* A_G, *and define* $e_H=eR_{G,H}$ *and* $A'_H=e_HA_He_H$, *for all subgroups H of G. Then the family* (A'_H) *satisfies conditions* 1.52(a), (b) *and so defines a subfunctor* $\mathbf{A}'$ *of* $\mathbf{A}$, *whose canonical filtration is* $(e_HA_{U,H}e_H)$.

Proof. Let H, K be subgroups of G such that $H\leqslant K$. Since $R_{G,H}$ is multiplicative, e_H is an idempotent; since A_H is associative, $A'_H=e_HA_He_H$ is a subalgebra of A_H. Also $e_H=e_KR_{K,H}$ so that

$$(e_H\alpha e_H)\,T_{H,K}=e_K(\alpha T_{H,K})e_K$$

for any α in A_H, by two applications of axiom G5. Verification of 1.52(b), and that $A'_{U,H}=e_HA_{U,H}e_H$ for all U,H, now present no difficulty.

Notation. We write $e\mathbf{A}e$ for the G-functor $\mathbf{A}'$ described in Lemma 3.37.

Proposition 3.38. *Let* $\mathbf{A}:\mathcal{S}(G)\to\mathcal{A}_k$ *be a multiplicative G-functor such that each algebra* A_H *is associative and has identity element* 1_H. *Let*

$$(2)\qquad 1_G=e_1+\ldots+e_n$$

be a decomposition of 1_G *as sum of mutually orthogonal idempotents* $e_1,\ldots,e_n$ *of* A_G. *Then*

$$(\mathrm{a})\qquad \mathfrak{D}(\mathbf{A})=\bigcup_{i=1}^{n}\mathfrak{D}(e_i\mathbf{A}e_i)\,,$$

and if e_i, e_j *are associated in* A_G, *then* $\mathfrak{D}(e_i\mathbf{A}e_i)=\mathfrak{D}(e_j\mathbf{A}e_j)$ $(i,j=1,\ldots,n)$.

(b) *If for each i* $(i=1,\ldots,n)$ e_i *is completely primitive in* A_G, *and if* D_i *is a defect group of* e_i, *then*

$$\mathfrak{D}(\mathbf{A})=j\{D_1,\ldots,D_n\}\,.$$

Proof. (a) follows directly from 3.37, 3.33 and 3.35. In (b), the existence of the defect groups D_i comes from 3.36, and the last equation is from (a), and the fact that $\bigcup_{i=1}^{n}j\{D_i\}=j\{D_1,\ldots,D_n\}$.

Remarks. 1) If e_i, e_j are associated, then we can take $D_i=D_j$ in (b). 2) If A_G satisfies the minimum condition for right ideals (and in particular, if A_G is finite-dimensional over a field k) then there always exists a decomposition (2) in which all the e_i are completely primitive in A_G. See also § 4.3.

3.4. Let A be an associative k-algebra with identity element 1, and $(A_{\mathfrak{U}})$ a G-family on A. We show in this paragraph that, if k is a Noetherian domain of finite dimension, it is possible to reduce the problem of calculating $\mathfrak{D}(A)$ to the case where k is a field; this is done by the usual arithmetic "localization". If A is also finitely generated as k-module, this means that we can reduce to the case where A is a finite-dimensional algebra over a field, and then apply the methods of §3.3 to "localize" still further, to the case of a completely primary algebra eAe (i.e. e completely primitive in A).

For the moment let k be any commutative ring with identity 1_k. Let $r, \phi : k \to r$ be as given in §1.6, and let $\Phi : \mathscr{A}_k \to \mathscr{A}_r$ be the functor described there. In contexts where it is not necessary to mention ϕ, Φ explicitly, we write $P\Phi = r \otimes_k P$, for any k-algebra P.

Proposition 3.41. (a) *With the notation above, $(r \otimes_k A_{\mathfrak{U}})$ is a G-family on $r \otimes_k A$, and $\mathfrak{D}(A) \supseteq \mathfrak{D}(r \otimes_k A)$.*

(b) ([4], Lemma 1) *If $k \subseteq r$ and $\phi : k \to r$ is the inclusion map, and if r has a k-basis $\{r_\mu\}$ containing $r_0 = 1_r$ as one of its elements, then $\mathfrak{D}(A) = \mathfrak{D}(r \otimes_k A)$.*

Proof. (a) It is trivial that $(r \otimes_k A_{\mathfrak{U}})$ is a G-family on $r \otimes_k A$. The other assertion follows from the fact that, for any $\mathfrak{U} \in \mathfrak{s}(G), A_{\mathfrak{U}} = A$ implies $r \otimes_k A_{\mathfrak{U}} = r \otimes_k A$.

(b) By (a), it is enough to prove that $\mathfrak{D}(A) \subseteq \mathfrak{D}(r \otimes_k A)$. Let $\mathfrak{U} = \mathfrak{D}(r \otimes_k A)$. Then $1_r \otimes_k 1$, the identity element of $r \otimes_k A$, lies in $r \otimes_k A_{\mathfrak{U}}$, i.e.

$$(1) \qquad 1_r \otimes_k 1 = \sum_\mu r_\mu \otimes_k \alpha_\mu \,,$$

for some α_μ in $A_{\mathfrak{U}}$. But since r is the direct sum of the k-modules $r_\mu k$, therefore $r \otimes_k A_{\mathfrak{U}}$ is the direct sum of the $r_\mu \otimes_k A_{\mathfrak{U}}$. Comparing the coefficients of r_0 in (1), we find $1 = \alpha_0 \in A_{\mathfrak{U}}$. Hence A is $\mathfrak{U}$-projective, and since $\mathfrak{U}$ is j-closed, this gives $\mathfrak{D}(A) \subseteq \mathfrak{U}$, and Proposition 3.41 is proved.

Remark. Proposition 3.41(b) shows that if r is an extension field of a field k, then $\mathfrak{D}(A) = \mathfrak{D}(r \otimes_k A)$.

Proposition 3.42. *Let A, $(A_{\mathfrak{U}})$ be as before, and assume that k is an integral domain. Let K be the field of fractions of k. Then*

(a) *If $\mathfrak{D}_0 = \mathfrak{D}(K \otimes_k A)$, there exists $v \in k$, $v \neq 0$, such that $v1 \in A_{\mathfrak{D}_0}$.*

(b) *If k is Noetherian, and if $\mathfrak{p}_1, \ldots, \mathfrak{p}_m$ are the minimal prime ideals belonging to the principal ideal (v) in k, then*

$$\mathfrak{D}(A) = \mathfrak{D}_0 \cup \mathfrak{D}_1 \cup \ldots \cup \mathfrak{D}_m \,,$$

where $\mathfrak{D}_i = \mathfrak{D}((k/\mathfrak{p}_i) \otimes_k A)$, for $i = 1, \ldots, m$.

Proof. (a) Since $K \otimes_k A$ is $\mathfrak{D}_0$-projective, $1_K \otimes_k 1 \in K \otimes_k A_{\mathfrak{D}_0}$, so that $1_K \otimes_k 1 =$

$\sum_j (u_j/v_j) \otimes_k \alpha_j$ (finite sum), for certain u_j, v_j ($v_j \neq 0$) in k and α_j in $A_{\mathfrak{X}_0}$. Multiplying by the product v of the v_j, and identifying $k \otimes_k A$ with A in the usual way, we have $v1 = \sum u_j(vv_j^{-1})\alpha_j \in A_{\mathfrak{X}_0}$.

(b) (For properties of the set of prime ideals belonging to an ideal in a Noetherian ring, see e.g. [2] Chapters 4 and 7.) The intersection $\mathfrak{p}_1 \cap \ldots \cap \mathfrak{p}_m$ is the "radical" $r((v))$ of (v) (in the sense of ideal theory, see [2] p. 8); moreover there is some positive integer s such that $r((v))^s \leqslant (v)$ ([2] p. 83). Therefore $(\mathfrak{p}_1 \ldots \mathfrak{p}_m)^s \leqslant (v)$.

For each $\mathfrak{p}_i$, we identify $(k/\mathfrak{p}_i) \otimes_k A$ with $A/\mathfrak{p}_i A$. To say that $(k/\mathfrak{p}_i) \otimes_k A$ is $\mathfrak{X}_i$-projective, is to say that $A_{\mathfrak{X}_i} \equiv A \bmod \mathfrak{p}_i A$. So there exists α_i in $A_{\mathfrak{X}_i}$ such that $1 + \alpha_i \in \mathfrak{p}_i A$. Therefore

$$(1) \qquad \prod_{i=1}^{m} (1+\alpha_i)^s \in (\mathfrak{p}_1 \ldots \mathfrak{p}_m)^s A \leqslant vA \leqslant A_{\mathfrak{X}_0},$$

by (a). Let $\mathfrak{X} = \mathfrak{X}_0 \cup \mathfrak{X}_1 \cup \ldots \cup \mathfrak{X}_m$. Then $A_{\mathfrak{X}} \geqslant A_{\mathfrak{X}_0} + A_{\mathfrak{X}_1} + \ldots + A_{\mathfrak{X}_m}$, and if we take (1) modulo $A_{\mathfrak{X}}$, we find that $1 \in A_{\mathfrak{X}}$. Therefore A is $\mathfrak{X}$-projective, and since $\mathfrak{X}$ is j-closed, $\mathfrak{X}(A) \subseteq \mathfrak{X}$. But $\mathfrak{X}(A)$ contains $\mathfrak{X}_0, \ldots, \mathfrak{X}_m$, by Proposition 3.41(a), hence $\mathfrak{X}(A) \supseteq \mathfrak{X}$. We have now $\mathfrak{X}(A) = \mathfrak{X}$, which proves Proposition 3.42.

Since $v \neq 0$, each of the prime ideals $\mathfrak{p}_i$ is non-zero. So if the dimension dim k (see [2] Chapter 11) is $\leqslant 1$, each $\mathfrak{p}_i$ is maximal, and $k/\mathfrak{p}_i$ is a field. This happens, in particular, if k is a Dedekind domain. More generally we have

Proposition 3.43. *Let A, $(A_{\mathfrak{U}})$ be as in Proposition 3.42, and suppose that k is a Noetherian domain of finite dimension. Then there exists a finite index set N, and for each $\nu \in N$ a field K_ν and a ring-homomorphism $\phi_\nu : k \to K_\nu$, such that*

$$\mathfrak{X}(A) = \bigcup_{\nu \in N} \mathfrak{X}(K_\nu \underset{k}{\otimes} A).$$

Proof. This is by induction on $d = \dim k$. For $d = 0$ (i.e. k a field) and $d = 1$, the statement is true, as we have seen. Suppose then $d > 1$ and the statement proved for dimensions $< d$. Since $v \neq 0$ (v the element appearing in Proposition 3.42), the $\mathfrak{p}_i$ all have "height 1", i.e. each $k/\mathfrak{p}_i$ is a Noetherian domain of dimension $d-1$ (Krull's principal ideal theorem, [2] p. 122). By induction hypothesis, there exists for each $i = 1, \ldots, m$ a finite index set N_i, and for each $\nu \in N_i$ a field K_ν and a ring-homomorphism $\phi_{\nu,i} : k/\mathfrak{p}_i \to K_\nu$, such that

$$(2) \qquad \mathfrak{X}_i = \mathfrak{X}((k/\mathfrak{p}_i) \underset{k}{\otimes} A) = \bigcup_{\nu \in N_i} \mathfrak{X}(K_\nu \underset{k/\mathfrak{p}_i}{\otimes} ((k/\mathfrak{p}_i) \underset{k}{\otimes} A)), \quad i = 1, \ldots, m.$$

Take N to be the union of the sets $\{0\}, N_1, \ldots, N_m$, which we may assume disjoint. For $\nu = 0$, we define $K_0 = K$ and $\phi_0 : k \to K$, the inclusion. For $\nu \in N_i$, define K_ν as

above, and ϕ_ν to be the composite map $\phi_\nu : k \xrightarrow{\text{nat}} k/\mathfrak{p}_i \xrightarrow{\phi_{\nu,i}} K$. There is a natural isomorphism

$$K_\nu \underset{\phi_\nu}{\otimes} A \cong K_\nu \underset{\phi_{\nu,i}}{\otimes} ((k/\mathfrak{p}_i) \underset{\text{nat}}{\otimes} A)\,,$$

where we have indicated explicitly the maps used to make the tensor products. Using these natural isomorphisms, we get the assertion of Proposition 3.43 by combining (2) with Proposition 3.42(b).

Propositions 3.41, 3.42, 3.43 have immediate applications to G-functors. If $\mathbf{A}: \mathcal{S}(G) \to \mathcal{A}_k$ is a G-functor with canonical filtration $(A_{\mathfrak{U},H})$, then the G-functor $r \otimes_k \mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_r$ has canonical filtration $(r \otimes_k A_{\mathfrak{U},H})$; so, for example, we deduce from 3.41(a) that $\mathfrak{D}(\mathbf{A}) \supseteq \mathfrak{D}(r \otimes_k \mathbf{A})$. We leave to the reader the task of writing out corresponding versions of 3.41(b), 3.42, 3.43.

3.5. *Injectivity*

In this paragraph $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_k$ is a fixed G-functor, such that A_G has identity element 1_G. The next definition gives a partial dual to that of projectivity (§3.1).

Definition. If $\mathfrak{U}$ is a set of subgroups of G, we say that A_G is $\mathfrak{U}$-*injective* if the map $R_{G,\mathfrak{U}} : \alpha \mapsto \oplus\, \alpha R_{G,U}$ is injective. Here $R_{G,\mathfrak{U}}$ is a map of A_G into the external direct sum $\oplus_{U \in \mathfrak{U}} A_U$.

Proposition 3.51. *If A_G is $\mathfrak{U}$-projective, then it is $\mathfrak{U}$-injective.*

Proof. If A_G is $\mathfrak{U}$-protective, then $1_G \in A_{\mathfrak{U},G}$, i.e. $1_G = \Sigma\, \alpha_U T_{U,G}$ (α_U are in A_U, sum over all $U \in \mathfrak{U}$). If now $\beta \in \operatorname{Ker} R_{G,\mathfrak{U}}$, we have by Axiom G5 $\beta = \beta \cdot 1_G = \Sigma\, \beta \cdot \alpha_U T_{U,G} = \Sigma\, (\beta R_{G,U} \cdot \alpha_U)\, T_{U,G} = 0$, since $\beta R_{G,U} = 0$ for all $U \in \mathfrak{U}$. Therefore A_G is $\mathfrak{U}$-injective.

Proposition 3.52. *The statements of Lemma* 3.11(a), (b), (c) *hold, with "projective" replaced by "injective".*

Proof. This is an easy exercise for the reader.

The next Proposition displays the duality of injectivity and projectivity very clearly.

Proposition 3.53. *Let $\mathbf{A}$, $\mathbf{A}' : \mathcal{S}(G) \to \mathcal{A}_k$ be G-functors, let $\theta : \mathbf{A} \to \mathbf{A}'$ be a morphism of G-functors (§1.5), and let $\mathfrak{U}$, $\mathfrak{V}$ be sets of subgroups of G. Then*

(a) *If A_G is $\mathfrak{U}$-injective, and if $\theta_U : A_U \to A'_U$ is injective for all $U \in \mathfrak{U}$, then $\theta_G : A_G \to A'_G$ is injective.*

(b) *If A'_G is $\mathfrak{V}$-projective, and if $\theta_V : A_V \to A'_V$ is surjective for all $V \in \mathfrak{V}$, then $\theta_G : A_G \to A'_G$ is surjective.*

Proof. (a) Suppose that θ_U is injective, for all $U \in \mathfrak{U}$; suppose also that $\alpha \in \operatorname{Ker} \theta_G$. Then $\alpha R_{G,U} \theta_U = \alpha \theta_G R'_{G,U} = 0$, hence $\alpha R_{G,U} = 0$, for all $U \in \mathfrak{U}$. Because A_G is $\mathfrak{U}$-injective, this implies $\alpha = 0$. Therefore θ_G is an injective map.

(b) Suppose that θ_V is surjective, for all $V \in \mathfrak{V}$; suppose also that $\beta \in A'_G$. Since A'_G is $\mathfrak{V}$-projective, there exists $\beta_V \in A'_V$ ($V \in \mathfrak{V}$) such that $\beta = \Sigma\, \beta_V T'_{V,G}$. There exists, for each $V \in \mathfrak{V}$, an element $\alpha_V \in A_V$ such that $\beta_V = \alpha_V \theta_V$ ($V \in \mathfrak{V}$). Therefore $\beta = \Sigma\, \alpha_V \theta_V T'_{V,G} = (\Sigma\, \alpha_V T_{V,G}) \theta_G \in \operatorname{Im} \theta_G$. This proves that θ_G is surjective.

Corollary 3.54. *Let $\mathbf{A}, \mathbf{A}' : \mathcal{S}(G) \to \mathcal{A}_k$ be G-functors such that A_G, A'_G have identity elements $1_G, 1'_G$ respectively. Let $\theta : \mathbf{A} \to \mathbf{A}'$ be a morphism of G-functors such that $1_G \theta_G = 1'_G$. Let $\mathfrak{U}$, $\mathfrak{V}$ be sets of subgroups of G such that A_G is $\mathfrak{U}$-injective and $\mathfrak{V}$-projective, and suppose that θ_U is injective for all $U \in \mathfrak{U}$, and that θ_V is surjective for all $V \in \mathfrak{V}$. Then $\theta_G : A_G \to A'_G$ is an isomorphism.*

Proof. θ_G is injective by Proposition 3.53(a). Therefore we have only to prove that θ_G is surjective, and by Proposition 3.53(b) this will follow if we show that A'_G is $\mathfrak{V}$-projective. Since A_G is $\mathfrak{V}$-projective, $1_G = \Sigma\, \alpha_V T_{V,G}$ for some $\alpha_V \in A_V$ ($V \in \mathfrak{V}$). Apply θ_G to this equation. We get $1'_G = \Sigma\, \alpha_V T_{V,G} \theta_G = \Sigma\, (\alpha_V \theta_V) T'_{V,G}$, and the last element is in $A'_{\mathfrak{V},G}$. Therefore A'_G is $\mathfrak{V}$-projective, and the corollary is proved.

§ 4. The transfer theorem

4.1. In §4.2 we shall prove the "transfer theorem" (Theorem 2) which shows how, for any G-functor $\mathbf{A}$, and for a given subgroup H of G, the maps $T_{H,G}$, $R_{G,H}$ induce maps which are "nearly" isomorphisms of k-algebras, between certain quotients of the filtration ideals in A_H and A_G. There are no restrictions on $\mathbf{A}$; in particular we do not need the hypothesis $A_G \cdot A_G = A_G$ of Theorem 1. Thus Theorem 2 applies when $\mathbf{A}$ is a G-functor into $\mathcal{M}_k$. However our main applications (§4.3) are significant only when the multiplication in the algebras A_H, A_G is non-trivial.

For the rest of this section, $\mathbf{A}$: $\mathcal{S}(G) \to \mathcal{A}_k$ is a fixed G-functor, and D, H are subgroups of G with $D \leqslant H$. Let

$$\mathfrak{X} = \{D^g \cap D \mid g \in G \setminus H\},$$

and

$$\mathfrak{Y} = \{D^g \cap H \mid g \in G \setminus H\}.$$

Thus $\mathfrak{X}$ is a set of subgroups of D, and $\mathfrak{Y}$ is a set of subgroups of H.

Lemma 4.11. *If $\beta \in A_{D,H}$, then there exists $\lambda \in A_{\mathfrak{Y},H}$ such that*

$$\beta T_{H,G} R_{G,H} = \beta + \lambda.$$

Proof. Since $\beta \in A_{D,H} = A_D T_{D,H}$, there exists some $\alpha \in A_D$ for which $\beta = \alpha T_{D,H}$. Then $\beta T_{H,G} = \alpha T_{D,G}$, so by Axiom G4

$$\beta T_{H,G} R_{G,H} = \alpha T_{D,G} R_{G,H} = \sum_{g \in \Gamma} \alpha C_{D,g} R_{D^g, D^g \cap H} T_{D^g \cap H, H},$$

where Γ is a (D, H)-transversal of G. The term $g = 1$ in this sum is $\alpha T_{D,H} = \beta$, and the sum λ of the remaining terms lies in

$$\sum_{g \in G \setminus H} A_{D^g \cap H} T_{D^g \cap H, H} = A_{\mathfrak{Y},H}.$$

Lemma 4.12. $A_{D,G} R_{G,H} \leqslant A_{D,H} + A_{\mathfrak{Y},H}$.

Proof. By 2.21(e), $A_{D,G} R_{G,H} \leqslant A_{D:G:H,H}$, and the last term can be written as

$$\sum_{g \in H} A_{D^g \cap H,H} + \sum_{g \in G \setminus H} A_{D^g \cap H,H} = A_{D,H} + A_{\mathfrak{Y},H},$$

since for g in H, $A_{D^g \cap H,H} = A_{D^g,H} = A_{D,H}$.

Lemma 4.13. $A_{\mathfrak{X},G}R_{G,H} \leqslant A_{\mathfrak{Y},H}$.

Proof. Suppose $X \in \mathfrak{X}$, so that $X = D^g \cap D$ for some g in $G \setminus H$. If now x is any element of G, then at least one of gx, x lies in $G \setminus H$, hence $X^x \cap H = D^{gx} \cap D^x \cap H$ is contained in a member of $\mathfrak{Y}$. Therefore $\mathfrak{X} : G : H \leqslant_H \mathfrak{Y}$, and the lemma follows from 2.21(e), (b).

Lemma 4.14. $A_{D,H} \cdot A_{\mathfrak{Y},H} \leqslant A_{\mathfrak{X},H}$ and $A_{\mathfrak{Y},H} \cdot A_{D,H} \leqslant A_{\mathfrak{X},H}$.

Proof. Suppose $Y \in \mathfrak{Y}$, so that $Y = D^g \cap H$ for some g in $G \setminus H$. If now h is any element of H, then gh^{-1} lies in $G \setminus H$. Hence $D^h \cap Y \leqslant_H D \cap D^{gh^{-1}} \in \mathfrak{X}$, and the first assertion of the lemma follows from 2.21(c), (b). The second assertion is proved similarly.

4.2. Since $\mathfrak{X}$ is a set of subgroups of D, and $\mathfrak{X} \leqslant_H \mathfrak{Y}$,

$$A_{\mathfrak{X},H} \leqslant A_{D,H} \cap A_{\mathfrak{Y},H},$$

by 2.21(b). Write $Q = A_{D,H} \cap A_{\mathfrak{Y},H}$ and let

$$q : A_{D,H}/A_{\mathfrak{X},H} \to A_{D,H}/Q$$

by the surjective k-map induced by the inclusion $A_{\mathfrak{X},H} \leqslant Q$ just mentioned. Let

$$s : (A_{D,H} + A_{\mathfrak{Y},H})/A_{\mathfrak{Y},H} \to A_{D,H}/Q$$

be the natural isomorphism. Both q, s are, of course, *multiplicative* k-maps.

Theorem 2. *Let* $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_k$ *be a G-functor, and let* D, H, $\mathfrak{X}$, $\mathfrak{Y}$, Q, q, s *be as defined above. Then*

(i) $T_{H,G}$ *induces a surjective k-map*

$$t : A_{D,H}/A_{\mathfrak{X},H} \to A_{D,G}/A_{\mathfrak{X},G},$$

and $R_{G,H}$ *induces a surjective k-map*

$$r : A_{D,G}/A_{\mathfrak{X},G} \to (A_{D,H} + A_{\mathfrak{Y},H})/A_{\mathfrak{Y},H},$$

and

$$trs = q. \tag{4.21}$$

(ii) t, r *are both multiplicative, i.e. they are both epimorphisms of k-algebras in the usual sense.*

(iii) Ker t *lies in* $Q/A_{\mathfrak{X},H}$, *which is an ideal of* $A_{D,H}/A_{\mathfrak{X},H}$ *which is annihilated by multiplication on either side by* $A_{D,H}/A_{\mathfrak{X},H}$. *Therefore* Ker t *lies in the radical of* $A_{D,H}/A_{\mathfrak{X},H}$.

(iv) Ker r *lies in* $(Q/A_{\mathfrak{X},H})t$, *which is an ideal of* $A_{D,G}/A_{\mathfrak{X},G}$ *which is annihilated by multiplication on either side by* $A_{D,G}/A_{\mathfrak{X},G}$. *Therefore* Ker r *lies in the radical of* $A_{D,G}/A_{\mathfrak{X},G}$.

Proof. (i) t, r are well-defined, because $A_{\mathfrak{X},H}T_{H,G} = A_{\mathfrak{X},G}$ (2.21(d)) and $A_{\mathfrak{X},G}R_{G,H} \leqslant A_{\mathfrak{Y},H}$ (Lemma 4.13). Also t is surjective because $A_{D,H}T_{H,G} = A_{D,G}$. Next we prove (4.21). If $x = \beta + A_{\mathfrak{X},H}$ is any element of $A_{D,H}/A_{\mathfrak{X},H}$, then $\beta \in A_{D,H}$ and we apply Lemma 4.11 to get

$$xtr = \beta T_{H,G}R_{G,H} + A_{\mathfrak{Y},H} = \beta + A_{\mathfrak{Y},H},$$

hence $xtrs = \beta + Q = xq$. This proves (4.21). Since s is bijective, $tr = qs^{-1}$ is surjective, hence r is surjective. This completes the proof of (i).

(ii) Let $\beta, \beta' \in A_{D,H}$. By 4.11, 4.14,

$$\beta \cdot (\beta' T_{H,G}R_{G,H}) \equiv \beta \cdot \beta' \mod A_{\mathfrak{X},H}.$$

Apply $T_{H,G}$ to both sides of this, and use Axiom G5. We get

$$\beta T_{H,G} \cdot \beta' T_{H,G} \equiv (\beta \cdot \beta')T_{H,G} \mod A_{\mathfrak{X},G}. \tag{1}$$

This shows that t is multiplicative. Now using Lemma 4.11, with β of that lemma replaced by $\beta \cdot \beta'$, β, β' in turn, we find

$$(\beta \cdot \beta')T_{H,G}R_{G,H} \equiv \beta \cdot \beta' \equiv \beta T_{H,G}R_{G,H} \cdot \beta' T_{H,G}R_{G,H} \mod A_{\mathfrak{Y},H}.$$

On the other hand, if we apply $R_{G,H}$ to (1) and use Lemma 4.13,

$$(\beta \cdot \beta')T_{H,G}R_{G,H} \equiv (\beta T_{H,G} \cdot \beta' T_{H,G})R_{G,H} \mod A_{\mathfrak{Y},H}.$$

Now let γ, γ' be any two elements of $A_{D,G}$. There exist β, β' in $A_{D,H}$ such that $\gamma = \beta T_{H,G}$, $\gamma' = \beta' T_{H,G}$. Putting these into the congruences just found gives

$$(\gamma \cdot \gamma')R_{G,H} \equiv \gamma R_{G,H} \cdot \gamma' R_{G,H} \mod A_{\mathfrak{Y},H}.$$

It follows that r is multiplicative.

(iii) Eq. (4.21) implies that Ker $t \leqslant$ Ker q, and by the definition of q, Ker $q = Q/A_{\mathfrak{X},H}$. The remaining statements of (iii) follow from Lemma 4.14.

(iv) Suppose that $y \in$ Ker r, and that x is any element of $A_{D,H}/A_{\mathfrak{X},H}$ such that $xt = y$. By (4.21), $xq = xtrs = yrs = 0$, i.e. $x \in Q/A_{\mathfrak{X},H}$. Therefore Ker $r \leqslant (Q/A_{\mathfrak{X},H})t$.

The other statements of (iv) follow from (iii) and the fact that t is surjective.

Remarks. 1) If $Q = A_{\mathfrak{X},H}$, then by (iii), (iv) both t, r are isomorphisms – as also is q, trivially. In particular this happens if D is a Sylow p-subgroup of G (for some prime p) and $H = N_G(D)$. For in this case $\mathfrak{X} = \mathfrak{Y}$).

2) Unless $H \geqslant N_G(D)$, D is a member of $\mathfrak{X}$, and $A_{D,H} = A_{\mathfrak{X},H}$, $A_{D,G} = A_{\mathfrak{X},G}$, so that the statements of Theorem 2 are all trivial. For this reason, nothing of value is lost if we assume that $H \geqslant N_G(D)$.

3) The maps t, r and q are all "near-isomorphisms"; where we say that a multiplicative epimorphism $\theta : A \twoheadrightarrow B$ (A, B k-algebras) is a near-isomorphism if $A \cdot \operatorname{Ker} \theta = \operatorname{Ker} \theta \cdot A = 0$. For the applications in §4.3, such maps are "as good as" isomorphisms, by the following lemma.

Lemma 4.22. *Let $\theta : A \twoheadrightarrow B$ be a near-isomorphism of associative k-algebras A, B. Let $E(A)$, $E(B)$ be the sets of idempotents in A, B respectively. Then*

(a) θ induces a bijection $E(A) \to E(B)$.

(b) If $e \in E(A)$, then e is completely primitive in A if and only if $e\theta$ is completely primitive in B.

(c) Elements $e, f \in E(A)$ are associated in A if and only if $e\theta$, $f\theta$ are associated in B.

Proof. (a) is an elementary consequence of the fact that for any $a \in A$, $n \in \operatorname{Ker} \theta$, we have $(a + n)^2 = a^2$. We leave the details to the reader. (b) See Remark 3, §4.3, below. (c) This follows from [16], p. 53, Prop. 1 and the fact that $\operatorname{Ker} \theta \leqslant \operatorname{rad} A$.

4.3. In this paragraph we apply Theorem 2 to a G-functor $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_k$ which satisfies the following.

Hypothesis 4.31. $\mathbf{A}$ is multiplicative, and for each subgroup H of G

(a) A_H is an associative k-algebra with identity element 1_H.

(b) $1_G R_{G,H} = 1_H$.

(c) 1_H has at least one *primitive decomposition in* A_H, viz.

$$1_H = f_1 + \ldots + f_n\,,$$

where $f_1, \ldots, f_n$ are mutually orthogonal idempotents which are completely primitive in A_H.

Remarks. 1) The last condition is assured if A_H is finitely-generated as k-module, and if k is a complete local principal ideal domain.

2) Let A be any associative k-algebra with identity element 1 which admits a primitive decomposition $1 = f_1 + \ldots + f_n$ in A. It is shown in [16] pp. 58, 59, that the Krull-Schmidt Theorem holds in the form: if $1 = f_1' + \ldots + f_{n'}'$ is another primitive

decomposition of 1 in A (we recall that, by our definition, the f_j, f_j' are *completely* primitive), then $n = n'$ and the f_j' can be so numbered that, for a suitable invertible element u of A, $u^{-1}f_j u = f_j'$ ($j = 1, \ldots, n$). Moreover, a small modification of the arguments indicated in [16], shows that every idempotent f of A can be expressed as the sum of a set (possibly empty) of mutually orthogonal primitive idempotents of A – which then must all lie in fAf. In particular, every primitive idempotent in A is completely primitiv in A.

3) Let $\theta : A \to B$ be a multiplicative epimorphism between the associative k-algebras A, B, and let e be a completely primitive idempotent in A. Then either $e\theta = 0$, or $e\theta$ is completely primitive in B. For θ induces an epimorphism $\bar{\theta} : eAe/\mathrm{rad}(eAe) \to (e\theta)B(e\theta)/\mathrm{rad}((e\theta)B(e\theta))$. Either $\bar{\theta} = 0$, in which case $e\theta = 0$, or $\bar{\theta}$ is an isomorphism, in which case $e\theta$ is completely primitive in B. When Ker θ lies in rad A, as happens for example when θ is a "near-isomorphism", then the second case must hold.

4) As special case of 3), we see that if S is an ideal of A, either e lies in S, or $e+S$ is completely primitive in A/S.

Now suppose that $\mathbf{A} : \mathcal{S}(G) \to \mathscr{A}_k$ satisfies Hypothesis 4.31. Let

$$(1) \qquad 1_G = e_1 + \ldots + e_n\,,$$

be any primitive decomposition of 1_G in A_G, and put $J = \{1, \ldots, n\}$. By Propositions 3.36, 3.37, each e_j has a defect group D_j, which is a subgroup of G determined up to conjugacy in G by

$$(2) \qquad \mathfrak{T}(e_j A_G e_j) = i_G\{D_j\}\,.$$

Lemma 4.32. *For given j in J, and any set $\mathfrak{U}$ of subgroups of G, e_j is in $A_{\mathfrak{U},G}$ if and only if $D_j \leqslant_G U$, for some U in $\mathfrak{U}$.*

Proof. By the definition of a defect base, $e_j \in A_{\mathfrak{U},G} \iff e_j A_G e_j$ is $\mathfrak{U}$-projective $\iff i_G\{D_j\} \subseteq \mathfrak{U}$. The lemma follows.

Let D be a fixed subgroup of G. By taking $\mathfrak{U} = \{D\}$ in the last lemma we have at once

Lemma 4.33. *For given j in J, $e_j \in A_{D,G}$ if and only if $D_j \leqslant_G D$.*

We write $D_j =_G D$ to mean that D_j, D are conjugate in G, and also $D_j <_G D$ to mean that $D_j \leqslant_G D$ and $D_j \neq_G D$. Let

$$J_0 = \{j \in J \mid D_j =_G D\}\,,\quad J_1 = \{j \in J \mid D_j <_G D\}\,,$$

and

$$J_2 = \{j \in J \mid D_j \nleqslant_G D\}\,.$$

Then J is the union of these three mutually disjoint sets (some of which may be empty). According to Lemma 4.33, $\{e_j \mid j \in J_0 \cup J_1\}$ is the set of e_j in the decomposition (1) which lie in $A_{D,G}$. The set $\{e_j \mid j \in J_0\}$ we call *the D-section of the decomposition (1)*; it is the set of those e_j whose defect group is conjugate in G to the given subgroup D.

Now take any subgroup H of G such that $H \geqslant N_G(D)$, and apply the map $R_{G,H}$ (which is multiplicative, by assumption) to (1):

$$1_H = 1_G R_{G,H} = \sum_{j \in J} e_j R_{G,H} ; \tag{3}$$

For each j in J, choose a suitable index set K_j (possibly empty) and a primitive decomposition in A_H of the idempotent $e_j R_{G,H}$

$$e_j R_{G,H} = \sum_{i \in K_j} f_{j,i} \quad (f_{j,i} \text{ primitive in } A_H) . \tag{4}$$

Putting (3), (4) together gives a primitive decomposition of 1_H in A_H:

$$1_H = \sum_{j \in J} \sum_{i \in K_j} f_{j,i} . \tag{5}$$

Each $f_{j,i}$ has a defect group $D_{j,i}$, which is a subgroup of H determined up to conjugacy in H by

$$\mathfrak{T}(f_{j,i} A_H f_{j,i}) = i_H(D_{j,i}) . \tag{6}$$

Here $\mathfrak{T}(f_{j,i} A_H f_{j,i})$ is the defect base relative to the H-family of ideals $(f_{j,i} A_{\mathfrak{U},H} f_{j,i})$ indexed by $\mathfrak{U}$ in $\mathfrak{s}(H)$ – it is therefore the defect base of the H-functor $f_{j,i}({}_H\mathbf{A}) f_{j,i}$, where ${}_H\mathbf{A}$ is the H-functor obtained from $\mathbf{A}$ by restriction to H (§1.7).

Since D is a subgroup of H, we may define the D-section of (5), viz. the set of those $f_{j,i}$ for which $D_{j,i} =_H D$. The main feature of our final Proposition 4.34 is that if $A_{D,G} = A_G$, which means that A_G is D-projective, then the D-sections of (1) and of (5) are in bijective correspondence.

Proposition 4.34. *Let $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_k$ be a G-functor satisfying Hypothesis 4.31, and let (1) be a primitive decomposition of 1_G in A_G. Let D be a fixed subgroup of G, and define the sets J_0, J_1, J_2 as above. Let H be a subgroup of G which contains $N_G(D)$, and define the primitive decompositions (4), (5) in A_H. Finally let $\mathfrak{X}$, $\mathfrak{Y}$ be the sets of subgroups defined in §4.1. Then*

(a) *For each j in J_0 there is exactly one element i in K_j such that $D_{j,i} =_H D$. Arranging notation so that this $i = 0$, we have*

$$e_j R_{G,H} \equiv f_{j,0} \bmod A_{\mathfrak{Y},H} , \tag{7}$$

and $f_{j,i} \in A_{\mathfrak{Y},H}$ *for all* $i \in K_j$, $i \neq 0$.
Moreover

$$(8) \qquad f_{j,0}T_{H,G} \equiv e_j \bmod A_{\mathfrak{X},G}.$$

(b) *The correspondence*

$$(9) \qquad e_j \leftrightarrow f_{j,0} \quad (j \in J_0)$$

defines a bijection between the D-section $\{e_j \mid j \in J_0\}$ *of (1) and the set* $\{f_{j,i} \mid j \in J_0 \cup J_1,\ i \in K_j,\ D_{j,i} =_H D\}$. *Therefore if* A_G *is D-projective, i.e. if* $J_2 = \emptyset$, *then (9) defines a bijection between the D-section of (1) and the D-section of (5).*

(c) *If* $j, k \in J_0$ *then* e_j, e_k *are associated in* A_G *if and only if* $f_{j,0}, f_{k,0}$ *are associated in* A_H.

Proof. (a) Since $H \geqslant N_G(D)$, all the members of $\mathfrak{X}$ are proper subgroups of D. Therefore if $j \in J_0$, $e_j \in A_{D,G}$ and $e_j \notin A_{\mathfrak{X},G}$, and so $e_j + A_{\mathfrak{X},G}$ is a non-zero, hence primitive, idempotent of $A_{D,G}/A_{\mathfrak{X},G}$. The map r of Theorem 2 takes

$$e_j + A_{\mathfrak{X},G} \to e_jR_{G,H} + A_{\mathfrak{Y},H} = \sum_{i \in K_j} (f_{j,i} + A_{\mathfrak{Y},H}).$$

Since r is a near-isomorphism, this last sum is a primitive idempotent; therefore exactly one of the terms $f_{j,i} + A_{\mathfrak{Y},H}$ is non-zero. Take $i = 0$ for this term; we have $e_jR_{G,H} + A_{\mathfrak{Y},H} = f_{j,0} + A_{\mathfrak{Y},H}$, which is equivalent to (7). Since $e_jR_{G,H} \in A_{D,H} + A_{\mathfrak{Y},H}$ (Lemma 4.12), $f_{j,0} = f_{j,0}e_jR_{G,H} \in A_{D,H} + A_{\mathfrak{Y},H} = A_{\{D\}\cup\mathfrak{Y},H}$. By Lemma 4.32 *, $D_{j,0} \leqslant_H U$, for some $U \in \{D\} \cup \mathfrak{Y}$. But $f_{j,0} \notin A_{\mathfrak{Y},H}$, so that $D_{j,0} \leqslant_H D$, and $f_{j,0} \in A_{D,H}$. Therefore we have

$$e_j + A_{\mathfrak{X},G} \xrightarrow{r} f_{j,0} + A_{\mathfrak{Y},H} \xrightarrow{s} f_{j,0} + Q.$$

Theorem 2 gives $trs = q$, and, by Lemma 4.22, that t, r, s, q all induce bijections on idempotents. Since q takes $f_{j,0} + A_{\mathfrak{X},H} \to f_{j,0} + Q$, we deduce that t takes $f_{j,0} + A_{\mathfrak{X},H} \to e_j + A_{\mathfrak{X},G}$, which is equivalent to (8). To complete the proof of (1), we must show that $D_{j,0} =_H D$. We already know that $D_{j,0} \leqslant_H D$. If $D_{j,0} \neq_H D$, we can take $D_{j,0} = D^* < D$. Then $f_{j,0} \in A_{D^*,H}$, so that $f_{j,0}T_{H,G} \in A_{D^*,H}T_{H,G} = A_{D^*,G}$, and then by (8), $e_j \in A_{D^*,G} + A_{\mathfrak{X},G} = A_{\{D^*\}\cup\mathfrak{X},G}$. But then Lemma 4.22 gives the contradiction that $D = D_j \leqslant_H U$, where $U = D^*$ or $U \in \mathfrak{X}$.

(b) In view of (a), we have only to show that if $D_{j,i} \leqslant_H D$, for some $j \in J$, $i \in K_j$, then $j \notin J_1$. This follows from Lemma 4.35, below.

* In this application of Lemma 4.32, H takes the place of G.

(c) Since t is a near-isomorphism, Lemma 4.22(c) and equation (8) show that $e_j + A_{\mathfrak{X},G}$, $e_k + A_{\mathfrak{X},G}$ are associated in $A_{D,G}/A_{\mathfrak{X},G}$ if and only if $f_{j,0} + A_{\mathfrak{X},H}$, $f_{k,0} + A_{\mathfrak{X},H}$ are associated in $A_{D,H}/A_{\mathfrak{X},H}$. Thus it is enough to prove that if e, f are completely primitive idempotents in an algebra A, both lying inside an ideal I of A, and both outside an ideal S of A, $S \leqslant I$, then e, f are associated in A if and only if $e + S$, $f + S$ are associated in I/S. We may assume $I = A$, Since Jacobson's conditions 3.3(1) imply that the x, y which occur there must be in I. Evidently, if e, f are associated in A, then $e + S$, $f + S$ are associated in A/S. Suppose conversely that $e + S$, $f + S$ are associated in A/S, so that x, y exist in A such that 3.3(1) hold mod S. As in the proof of [16] Prop. 1, p. 53, we may assume that $exf = x$ and $fye = y$. Now $xy = e + s$ for some s in $S \cap eAe$. But $S \cap eAe$ is a proper ideal of eAe, since $e \notin S$, hence lies in rad (eAe), which is contained in rad A. Similarly, $yx = f + t$ for t in $\operatorname{rad}(fAf) \leqslant \operatorname{rad} A$. Now the argument in the proof of [16] Prop. 1, p. 53, shows that e, f are associated in A.

Lemma 4.35. *If $D_{j,i} =_H D$ for some $j \in J$, $i \in K_j$, then $D_j \geqslant_G D$.*

Proof. From (4) we have

$$f_{j,i} = e_j R_{G,H} \cdot f_{j,i} .$$

Now $e_j \in A_{D_j,G}$, so that $e_j R_{G,H} \in A_{D_j:G:H,H}$ by 2.21(e). By assumption, $f_{j,i} \in A_{D,H}$. Therefore 2.21(c) gives

$$f_{j,i} \in A_{D_j:G:H,H} \cdot A_{D,H} \leqslant A_{\mathfrak{U},H} ,$$

where $\mathfrak{U} = (D_j:G:H):H:D = \{(D_j^g \cap H)^h \cap D \mid g \in G, h \in H\}$.
By Lemma 4.32 this implies that $D = D_{j,i} \leqslant_H D_j^{gh} \cap D$, for some $g \in G$, $h \in H$. But $D_j^{gh} \cap D \leqslant_G D_j$, hence $D \leqslant_G D_j$.

Remarks. (1) Lemma 4.35 makes possible a sharpening of the statement of 4.34(b); the correspondence (9) will give a bijection between the D-sections of (1), (5) respectively, if D is *maximal* among the defect groups $D_j (j \in J)$, i.e. if there exists no $j \in J$ such that $D_j >_G D$.

(2) In general, there may well exist $j \in J_2$, $i \in K_j$ such that $D_{j,i} =_H D$. Then $f_{j,i}$ must lie in $A_{D,H}$, and $f_{j,i} + A_{\mathfrak{X},H}$ is a primitive idempotent of $A_{D,H}/A_{\mathfrak{X},H}$. Therefore t maps $f_{j,i} + A_{\mathfrak{X},H}$ to a primitive idempotent of $A_{D,H}/A_{\mathfrak{X},G}$, but this need not be any of the idempotents $e_k + A_{\mathfrak{X},G} (k \in J_0)$.

§ 5. Examples

In these examples G is a given finite group. We use the notation of § 1.8, and define each G-functor $\mathbf{A} = (A, T, R)$ in terms of the maps $T_{H,K}$, $R_{K,H}$, $C_{H,g}$. H, K are always subgroups of G, $H \leqslant K$; g is an element of G. The verifications of the Conditions 1.83(a)-(i) are not given. Z, Q and C denote the rings of rational integers, rational numbers, and complex numbers respectively. We observe that all the G-functors in § § 5.1–5.5 are *multiplicative*.

5.1. *Character rings*

We define the *character ring functor on* G to be the following G-functor $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_Z$: for each H, A_H is the character ring (ring of generalized characters, [6] p. 272) of H; $T_{H,K}$ maps $\psi \in A_H$ to the induced character $\psi^K \in A_K$; $R_{K,H}$ maps $\chi \in A_K$ to its restriction $\chi_H = \chi|_H \in A_H$; $C_{H,g}$ maps $\psi \in A_H$ to the character $\psi^g \in A_{H^g}$ defined by $\psi^g(u) = (gug^{-1})$ $(u \in H^g)$. Conditions 1.83(a)-(i) follow from well-known theorems (see [6] chapter VI).

The defect base $\mathfrak{D}(\mathbf{A})$ is the set $\mathfrak{E}$ of all elementary subgroups ([6] p. 284) of G: this is a classical theorem * of R. Brauer. The methods given in § § 3.3, 3.4 for calculating defect bases for general G-functors are based on the proofs of Roquette [22] and Brauer and Tate [4], and it is therefore not surprising that these ingredients can be reassembled, and in several ways, to prove that $\mathfrak{D}(\mathbf{A}) = \mathfrak{E}$ when $\mathbf{A}$ is the character ring functor on G. We sketch one such proof here.

Artin's theorem [6] p. 279, shows that

$$(1) \qquad |G|\, 1_G \in A_{\mathfrak{C},G}\,,$$

where $\mathfrak{C}$ is the set of all cyclic subgroups of G. Therefore $\mathfrak{C} \supseteq \mathfrak{D}(K_0 \otimes \mathbf{A}) = \mathfrak{D}(K_0 \otimes A_G)$, where $K_0 = Q$, the rational field. But if $\mathfrak{U} \in \mathfrak{s}(G)$ and $K_0 \otimes \mathbf{A}$ is $\mathfrak{U}$-projective, then it is $\mathfrak{U}$-injective (3.51), and this is easily seen to imply $i_G \mathfrak{U} \supseteq \mathfrak{C}$. Therefore $\mathfrak{D}(K_0 \otimes A_G) = \mathfrak{C}$.** Now (1) shows that we may take $v = |G|$ in Proposition 3.42(a), and so from 3.42(b) we deduce

$$(2) \qquad \mathfrak{D}(A) = \bigcup_{i=0}^{m} \mathfrak{D}(K_i \otimes A_G)\,,$$

where $p_1, \ldots, p_m$ are the prime divisors of $|G|$, and $K_i = Z/(p_i)$ $(i = 1, \ldots, m)$. By 3.41(b) we can replace each K_i in (2) by its algebraic closure $\bar{K}_i$. Take fixed $i \in \{1, \ldots, m\}$, and write $p = p_i$. Let $\{G_{p,\nu} | \nu \in N_i\}$ (N_i is an index set) be the

* See [6] chapter VI for literature on this and other "induction theorems". The usual proof of Brauer's theorem proves only that $\mathbf{A}$ is $\mathfrak{E}$-projective, i.e. $\mathfrak{E} \supseteq \mathfrak{D}(\mathbf{A})$. But it has been known for a long time that $\mathfrak{E} = \mathfrak{D}(\mathbf{A})$ (see e.g. [10]).

** This could be proved directly by using Proposition 3.41(b) to replace K_0 by the complex field.

set of the p-regular classes of G. For each $\nu \in N_i$, let $S_{p,\nu}$ be the "p-superclass" (Roquette [22]) consisting of all g in G whose p-regular part lies in $G_{p,\nu}$. Then $\bar{K}_i \otimes A_G$ can be identified with the algebra of all functions from G into $\bar{K}_i$ which are constant on each p-superclass $S_{p,\nu}(\nu \in N_i)$. Hence $1 \in \bar{K}_i \otimes A_G$ has primitive decomposition $1 = \Sigma_{\nu \in N_i} e_{p,\nu}$, where $e_{p,\nu}$ is the idempotent function which takes value 1 on $S_{p,\nu}$ and zero outside $S_{p,\nu}$. The classical argument of Brauer ([6] §40) can be adapted to prove that each defect *group* (§3.3) of $e_{p,\nu}$ has the form $D_{p,\nu} = \langle x_{p,\nu}, P_{p,\nu}\rangle$, where $x_{p,\nu} \in G_{p,\nu}$ and $P_{p,\nu}$ is a Sylow p-subgroup of $C_G(x_{p,\nu})$. Then 3.37 gives

$$\mathfrak{T}(K_i \otimes A_G) = \mathfrak{T}(\bar{K}_i \otimes A_G) = j_G\,\{D_{p_i,\nu} | \nu \in N_i\}\ .$$

Now we put this into (2), for $i = 1, ..., m$, and obtain

$$\mathfrak{T}(A) = \mathfrak{C} \cup j_G\{D_{p_i,\nu} \mid \nu \in N_i, i = 1, ..., m\} = \mathfrak{C}\quad .$$

5.2. *Relative Grothendieck rings*

Let r be a commutative ring, and H a fixed subgroup of G. Lam and Reiner [19, 20, 21] define the *relative Grothendieck ring* $a(G,H)$ to be the Abelian group generated by symbols $[M]$, one for each rG-isomorphism class of (right, unital, finitely-generated) rG-modules, with defining relations $[L] - [M] + [N] = 0$ for every short exact sequence $0 \to L \to M \to N \to 0$ of rG-modules which is rH-split. Multiplication is defined from $[L]\,[M] = [L \otimes_r M]$ in the usual way.

Now let Y be a fixed *normal* subgroup of G. We define the *Grothendieck ring functor for r, G relative to Y* to be the following G-functor $\mathbf{A} : \mathcal{S}(G) \to \mathcal{A}_Z$: for each H, $A_H = a(H, H \cap Y)$; $T_{H,K}$ takes $[L]$, L an rH-module, to $[L^K]$, where $L^K = L \otimes_{rH} rK$ is the induced module; $R_{K,H}$ takes $[M]$, M an rK-module, to $[M_H]$, where M_H is the restriction of M to H; $C_{H,g}$ takes $[L]$ to $[L^g]$, where L^g is the rH^g-module made from the rH-module L through the isomorphism $u \mapsto gug^{-1}$ $(u \in H^g)$. It is necessary to check the consistency of these definitions. For example, we must show that any $H \cap Y$-split short exact sequence of rH-modules is converted by induction to K into a $K \cap Y$-split short exact sequence of rK-modules. This can be done by a modification of the argument used by Lam and Reiner to prove [20] Prop. 2.1. Consistency of $R_{K,H}$, $C_{H,g}$ is easy to prove.

Special cases. If we take $r = C$, $Y = \{1\}$ we get a G functor isomorphic to the character ring functor on G. If r is any algebraically closed field, $Y = \{1\}$, then A_H is isomorphic to the ring of Brauer characters of H over r. For arbitrary r, and $Y = G$, A_H is the representation algebra of [13]. For further generalizations, see Dress [7, 8].

The transfer theorem [14] Theorem 1, is not a special case of Theorem 2 of this paper, because the filtration $(a_{\mathfrak{U}}(H))$ used in [14] is not canonical. $a_{\mathfrak{U}}(H)$ is spanned, not by all $[L]$ for which L is an rH-module induced from some rU-module, $U \in \mathfrak{U}$,

but by the direct summands of such modules; evidently $a_{\mathfrak{U}}(H)$ contains the canonical $A_{\mathfrak{U},H}$. The "transfer of modules" (Theorem 2 of [14]) is, however, deducible from a special case of our present Theorem 2, by means of Proposition 4.34, see §5.5.

5.3. *Cohomology*

Let A be a G-algebra over k (see §5.5). We define two *cohomology ring functors for G with coefficients in A*, namely the following G-functors $\mathbf{A}$, $\hat{\mathbf{A}}: \mathcal{S}(G) \to \mathcal{A}_k$: for each H, A_H is the ordinary cohomology ring $H^*(H, A)$, and $\hat{A}_H$ is the Tate cohomology ring $\hat{H}^*(H, A)$ ([5] Chapter XII); $T_{H,K}$ is the Eckmann transfer [9], $R_{K,H}$ the restriction map, and $C_{H,g}$ the conjugation map c_g, as defined in [5] pp. 254, 255 – these are for Tate cohomology, but apply also to ordinary cohomology. The same is true of the formulae appearing in [5] pp 255–258, which show that our conditions 1.83(a)–(i) hold for $\mathbf{A}$ and $\hat{\mathbf{A}}$; moreover these are both cohomological functors (§1.4). For the rest of this paragraph we confine ourselves to the functor $\mathbf{A}$; analogous results hold for $\hat{\mathbf{A}}$.

$A_H = H^*(H, A)$ has a grading, for each H. This grading is preserved by all the maps $T_{H,K}$, $R_{K,H}$, $C_{H,g}$. Write $A_H^n = H^n(H, A)$ $(n \geqslant 0)$. Since $A_H^m \cdot A_H^n \leqslant A_H^{m+n}$ for all $m, n \geqslant 0$, A_H^0 can be regarded as a k-subalgebra of A_H. The family $(A_H^0)_{H \leqslant G}$ defines a subfunctor $\mathbf{A}^0$ of $\mathbf{A}$, and it is easy to check that $\mathbf{A}^0$ is isomorphic to the G-functor $\mathbf{f}(A)$ defined in §5.5 *.

Proposition 5.31. $\mathfrak{T}(\mathbf{A}) = \mathfrak{T}(\mathbf{A}^0) = \mathfrak{T}(\mathbf{f}(A))$.

Proof. We have only to prove the first equality; the second comes from $\mathbf{A}^0 \cong \mathbf{f}(A)$. Let $\mathfrak{U} \in \mathfrak{s}(G)$. Then $\mathbf{A}$ is $\mathfrak{U}$-projective $\Longleftrightarrow 1_G \in A_{\mathfrak{U},G} \Longleftrightarrow 1_G \in A^0_{\mathfrak{U},G} \Longleftrightarrow \mathbf{A}^0$ is $\mathfrak{U}$-projective, since $1_G \in A_G^0$ and the maps $T_{U,G}$ respect grading. This completes the proof.

Theorem 2 gives directly, when applied to the cohomology ring functors, transfer theorems for cohomology. We look at one special case. Let D be a Sylow p-subgroup of G, for some prime p. Take $H = N_G(D)$, then $\mathfrak{X} = \mathfrak{Y}$ (see Remark 1, §4.2) and we have the isomorphism

$$(1) \qquad t: A_{D,H}/A_{\mathfrak{X},H} \to A_{D,G}/A_{\mathfrak{X},G}\,.$$

Let us also assume that k satisfies the

Hypothesis 5.32. k is a field of characteristic p, or k is a complete local principal ideal domain with maximal ideal m such that k/m has characteristic p.

* For Tate cohomology, $\hat{A}_H^0 \cong A_H/A_{\{1\},H}$ in the notation of §5.5.

By (3.12), every member of $\mathfrak{X}(\mathbf{A})$ is a p-subgroup of G. Hence A_G is $\{D\}$-projective, i.e. $A_{D,G} = A_G$. Similarly $A_{D,H} = A_H$. Taking components of fixed degree $n \geqslant 0$ in (1) we get

$$A_H^n / A_{\mathfrak{X},H}^n \cong A_G^n / A_{\mathfrak{X},G}^n,$$

from which D.L. Johnson's transfer theorem [17] is easily deduced.

5.4. *K-theory*

We define the K^* *functor on* G to be the following G-functor $\mathbf{A}' : \mathcal{S}(G) \to \mathcal{A}_Z$: for each H, $A'_H = K^*(B_H)$, where B_H is a classifying space for H, and K^* is the Atiyah-Hirzebruch functor described in [1]. If $H \leqslant K$, we take a classifying space B_K for K, then factorize the universal covering $E_K \to B_K$ as $E_K \to E_K/H \xrightarrow{f} B_K$, so that $B_H = E_K/H$ (space of H-orbits on E_K) is a classifying space for H. The map $f : B_H \to B_K$ is a finite covering, and we define $T'_{H,K} = f_!$, $R'_{K,H} = f^!$, as in [1]. Define $C'_{H,g} = c^!$, where $c : B_{H^g} \to B_H$ is induced by the isomorphism $H^g \to H$ which takes $u \mapsto gug^{-1}$ $(u \in H^g)$.

Atiyah [1] gives a connection between the ordinary cohomology ring functor for G with coefficients in Z, and the character ring functor on G, which we denote $\mathbf{A}$, as in 5.1. This is done through the intermediary of the K^* functor $\mathbf{A}'$ on G, and it is interesting * to describe the "functorial" part (i.e. the easy part!) of the argument [1] §§7–11, which connects $\mathbf{A}$ with $\mathbf{A}'$.

For each H, A_H has the topology defined by taking as system of neighbourhoods of zero the powers $(I_H^n)_{n \geqslant 1}$ of the "augmentation ideal" $I_H = \{\psi \in A_H \mid \psi(1) = 0\}$. It follows from [1] Lemma 6.7 that $T_{H,K}$, $R_{K,H}$, $C_{H,g}$ are all continuous **, and therefore if $\hat{A}_H$ denotes the completion of A_H with respect to this topology, they induce maps $\hat{T}_{H,K}$, etc. which make $\hat{\mathbf{A}} = (\hat{A}, \hat{T}, \hat{R})$ into a G-functor in such a way that the family of natural maps $A_H \to \hat{A}_H$ provides a surjective morphism $\mathbf{A} \to \hat{\mathbf{A}}$ of G-functors. By 3.32(a), $\mathfrak{X}(\mathbf{A}) \supseteq \mathfrak{X}(\hat{\mathbf{A}})$, and since $\mathfrak{X}(A) = \mathfrak{C}$ (see §5.1), $\hat{\mathbf{A}}$ is $\mathfrak{C}$-projective. It is not hard to see ([1] Lemma 8.3) that $\hat{\mathbf{A}}$, like $\mathbf{A}$, is $\mathfrak{C}$-injective.

There is a naturally defined morphism $\alpha : \mathbf{A} \to \mathbf{A}'$ which factorizes through $\mathbf{A} \to \hat{\mathbf{A}}$ to give a morphism $\hat{\alpha} : \hat{\mathbf{A}} \to \mathbf{A}'$ ([1] §7). Atiyah shows ([1] §8) that $\hat{\alpha}_C$ is injective for all $C \in \mathfrak{C}$, and ([1] §10) that $\hat{\alpha}_E$ is surjective for all $E \in \mathfrak{C}$. Since $\hat{\alpha}_G$ maps $\hat{1}_G$ to $1'_G$, we may apply Corollary 3.54 to see that $\alpha_G : \hat{A}_G \to A'_G$ is an isomorphism – this is the main part of [1] Theorem 7.2. It follows that $\hat{\alpha} : \hat{\mathbf{A}} \to \mathbf{A}'$ is an isomorphism of G-functors.

5.5. *G-algebras*

A G-algebra over k ([15] §4) is a k-algebra with identity element 1, on which G

* Lam gives a very similar discussion in [18] pp. 115–116.

** For general n, $I_H^n T_{H,K}$ is not contained in I_K^n, but Atiyah's Lemma 6.7 shows that for given $m \geqslant 1$, there exists $n = n(m) \geqslant 1$ such that $I_H^n T_{H,K} \leqslant I_K^m$.

acts as a group of k-algebra automorphisms: $g \in G$ acts on $\alpha \in A$ to give $\alpha^g \in A$, and this G-action makes A into a right kG-module, and $(\alpha\beta)^g = \alpha^g\beta^g$ for all $\alpha, \beta \in A$ and $g \in G$. We define the *G-functor* $\mathbf{f}(A)$ *associated with the G-algebra A* to be the following G-functor into $\mathscr{A}_k$: for each H, $A_H = \{\alpha \in A \mid \alpha^h = \alpha$, all $h \in H\}$; $T_{H,K}$ takes $\alpha \in A_H$ to $\Sigma\, \alpha^v$, summed over an H-transversal $\{v\}$ of K; $R_{K,H}$ is the inclusion map $(A_K \leqslant A_H)$; $C_{H,g}$ takes $\alpha \in A_H$ to $\alpha^g \in A_{Hg}$. Conditions 1.83(a)–(i) are easily checked (see [15] pp. 139, 140). Moreover $\mathbf{f}(A)$ *is a cohomological G-functor* (§1.4), for if $\beta \in A_K$, then $\beta R_{K,H} T_{H,K} = \Sigma\, \beta^v = (K : H)\beta$. It is evidently multiplicative. These functors associated with G-algebras provide our first example in which the k-algebras A_H need be neither commutative nor associative.

Particularly important examples of G-algebras occur (a) in the "vertex theory" of kG-modules, and (b) in Brauer's Theory of blocks of modular representations.

(5.5a) Let M be a kG-module. Make $A = \mathrm{End}_k M$ into a G-algebra by defining, for $\theta \in A$, $g \in G$, the k-map $\theta^g \in A$ to take each $m \in M$ to $((mg^{-1})\theta)g$. Following Dress [8] we say M is $\mathfrak{U}$-projective for given $\mathfrak{U} \in \mathfrak{s}(G)$, if $A_G = \mathrm{End}_{kG}(M)$ is $\mathfrak{U}$-projective, i.e. if $1 = 1_G \in A_{\mathfrak{U},G}$. The defect base $\mathfrak{T}(A_G) = \mathfrak{T}(\mathbf{f}(A))$ defines Dress's $\mathfrak{U}(M)$ ([8] p. 1034). If M is *completely indecomposable*, i.e. if 1_G is a completely primitive idempotent of A_G, then 1_G has a defect group D, a subgroup of D determined up to conjugacy in G by $\mathfrak{T}(A_G) = i_G\{D\}$; D is a *vertex* of M ([11, 12]; see [15] §5).

If M is finitely generated and if k satisfies Hypothesis 5.32, for some prime p, then $\mathbf{A} = \mathbf{f}(A)$ satisfies Hypothesis 4.31, and the argument in §4.3 can be interpreted as follows. Let

(1) $$M = M_1 \oplus \ldots \oplus M_n\,, \qquad (M_j \text{ indecomposable } kG\text{-modules}),$$

and for each $j \in J = \{1, \ldots, n\}$

(3) $$(M_j)_H = \bigoplus_{i \in K_j} M_{j,i}\,, \qquad (M_{j,i} \text{ indecomposable } kH\text{-modules}),$$

and so

(5) $$M_H = \bigoplus_{j,i} M_{j,i}$$

is a complete decomposition of M_H. Proposition 4.34 shows that for each M_j in (1) of vertex D, there is exactly one $M_{j,0}$ of vertex D among the $M_{j,i}$ $(i \in K_j)$, the other $M_{j,i}$ $(i \in K_j,\, i \neq 0)$ being all $\mathfrak{Y}$-projective. Also $(M_{j,0})^G \cong M_j \oplus$ (an $\mathfrak{X}$-projective kG-module). Finally, if M is D-projective, then $M_j \longleftrightarrow M_{j,0}$ is a bijective correspondence between the set of terms of vertex D in (1), and the set of terms of vertex D in (5), and two such terms M_j, M_k in (1) are isomorphic kG-modules if and only if $M_{j,0}, M_{k,0}$ are isomorphic kH-modules *. In particular if M is itself an indecompo-

* This last statement comes from 4.34(c), since for example $M_j = Me_j$, $M_k = Me_k$ are isomorphic kG-modules if and only if e_j, e_k are associated in A_G.

sable kG-module of vertex D, we may recover from this the module correspondence given in [14] Theorem 2.

(5.5b) The group algebra $A = kG$ becomes a G-algebra, defining α^g to be $g^{-1}\alpha g$ ($\alpha \in A, g \in G$). A_G is the centre of $A = kG$, and $A_{H,G}$ is spanned as k-module by the sums (C_i') of the H-classes C_i' in G (elements g, g' are in the same H-class if there exists $h \in H$ with $g = h^{-1}g'h$); in particular $A_G = A_{G,G}$ is spanned by the sums (C_j) of the G-classes (conjugacy classes) C_j of G.

If k satisfies Hypothesis 5.32 and if the polynomial $\lambda^{|G|} - 1$ factorises linearly over k, then each primitive idempotent e_j of A_G corresponds to a uniquely defined p-block $\mathfrak{B}_j$, and a defect group D_j of e_j is the same as a defect group of $\mathfrak{B}_j$ in Brauer's original usage (see, e.g. [3]). This connection is discussed in [15].

For the moment let A be any G-algebra over k, let D be a fixed p-subgroup of G, and $H = N_G(D)$. We use the notation of §§4.1, 4.2. If k is a field, it is easy to find k-bases for the ideals which appear in Theorem 2, whenever A is a G-algebra with a permutation base [15] §5, and to show that in this case $A_{D,H} \cap A_{\mathfrak{Y},H} = A_{\mathfrak{X},H}$; so by Remark 1 of §4.2, the maps t, r, q are all isomorphisms. Let $\mathfrak{Z}$ be the set of all proper subgroups of D. Since $H = N_G(D)$, $\mathfrak{X} \subseteq \mathfrak{Z}$. We may replace $\mathfrak{X}$ by $\mathfrak{Z}$ in Theorem 2, and we have an isomorphism

$$rs : A_{D,G}/A_{\mathfrak{Z},G} \to A_{D,H}/A_{\mathfrak{Z},H} \,.$$

In our case ($A = kG$, k a field satisfying Hypothesis 5.32) we find that for any $\mathfrak{U} \in \mathfrak{s}(H)$, $A_{\mathfrak{U},H}$ has as k-basis the set of all the (C_i') for which $\{D_i'\} \leqslant_H \mathfrak{U}$ (see [15] §5), hence that $A_{D,H}/A_{\mathfrak{Z},H}$ has k-basis $\{(C_i') + A_{\mathfrak{Z},H} | D_i' =_H D\}$; here D_i' is the "class-defect group" of C_i', defined up to conjugacy in H as a Sylow p-subgroup of $C_H(g), g \in C_i'$. We may define similarly ideals $B_{D,H}$, $B_{\mathfrak{Z},H}$ in the centre B_H of the H-algebra $B = kH$; $B_{D,H}/B_{\mathfrak{Z},H}$ has k-basis $\{(C_h'') + B_{\mathfrak{Z},H} \mid D_h'' =_H D\}$, where (C_h'') are the sums of elements in the H-classes of H. Now if an H-class C_i' has class-defect group D, then there is some $g \in C_i' \cap C_G(D)$, hence C_i' lies in H. Therefore the set of such C_i', coincides with the set of H-classes C_h'' of H which have class-defect group D; so the inclusion $B \leqslant A$ induces an isomorphism of k-algebras,

$$h : B_{D,H}/B_{\mathfrak{Z},H} \to A_{D,H}/A_{\mathfrak{Z},H} \,.$$

Hence we have the isomorphism

$$rsh^{-1} : A_{D,G}/A_{\mathfrak{Z},G} \to B_{D,H}/B_{\mathfrak{Z},H} \,.$$

This is equivalent to Brauer's homomorphism [3] in the present case; it gives a proof of Brauer's "first main theorem", because the set of blocks $\mathfrak{B}_j$ of G with defect group D is in bijective correspondence with the set of primitive idempotents of the commutative algebra $A_{D,G}/A_{\mathfrak{Z},G}$, and hence, by rsh^{-1}, with the set of blocks $\mathfrak{b}_j$ of H with defect group D.

5.6. *Free G-functors*

In this paragraph we outline a theory of a kind of "relatively free" G-functors. We confine ourselves, for simplicity, to functors into the module category $\mathfrak{M}_k$.

First we define a *G-semifunctor* $\mathbf{B} = (B, R)$ *over* k to be a contravariant functor $\mathbf{B} : \mathcal{S}_0(G) \to \mathfrak{M}_k$, where $\mathcal{S}_0(G)$ is the category with the same objects as $\mathcal{S}(G)$, but whose morphisms are restricted to the inclusions $(H, 1, K)$ in $\mathcal{S}(G)$. Thus to each H, B assigns a k-module B_H, and to each pair H, K $(H \leqslant K)$, R assigns a k-map $R_{K,H} : B_K \to B_H$ in such a way that conditions like 1.83(b) hold. A natural map between such G-semifunctors is called a *morphism of G-semifunctors over k*; we get then the *category* $\mathfrak{M}_k^0(G)$ *of G-semifunctors over k*. There is a "forgetful functor" from $\mathfrak{M}_k(G)$ into $\mathfrak{M}_k^0(G)$, which takes $\mathbf{A} : \mathcal{S}(G) \to \mathfrak{M}_k$, a G-functor, to $\mathbf{A}_0 : \mathcal{S}_0(G) \to \mathfrak{M}_k$, namely $\mathbf{A}$ regarded as G-semifunctor. We shall define a functor $\mathfrak{M}_k^0 \to \mathfrak{M}_k$ which is adjoint to the forgetful functor.

For each morphism $\pi = (H, g, K)$ in $\mathcal{S}(G)$, define a symbol $[H, g, K]$, in such a way that $[H, g, K] = [H', g', K']$ if and only if $H = H'$, $K = K'$ and $g^{-1}g' \in K$. Let $\mathbf{B} : \mathcal{S}_0(G) \to \mathfrak{M}_k$ be a given G-semifunctor. For each subgroup H of G, define A_H to be the direct sum $\oplus\, B_D * [D, x, H]$, summed over all distinct symbols $[D, x, H]$ with $D^x \leqslant H$, and where each term $B_D \otimes [D, x, H]$ is the set of symbols $\beta \otimes [D, x, H]$ $(\beta \in B_D)$, made into a k-module by the requirement that $\beta \mapsto \beta \otimes [D, x, H]$ be a k-map. Then we define the *relatively free G-functor* $\langle\mathbf{B}\rangle$ *generated by* $\mathbf{B}$ to be the following G-functor $\mathbf{A} = \langle\mathbf{B}\rangle : \mathcal{S}(G) \to \mathfrak{M}_k$: for each H, A_H is the k-module just described; the maps $T_{H,K}$, $R_{H,F}$, $C_{H,g}$ $(K \geqslant H \geqslant F,\ g \in G)$ are defined by

$$T_{H,K} : \beta \otimes [D, x, H] \mapsto \beta \otimes [D, x, K]\ ,$$

$$R_{H,F} : \beta \otimes [D, x, H] \mapsto \Sigma\, \beta R_{D, D \cap F^{x^{-1}g^{-1}}} \otimes [D \cap F^{x^{-1}g^{-1}}, gx, F]\ ,$$

summed over g in a $(D, F^{x^{-1}})$-transversal of $H^{x^{-1}}$,

$$C_{H,g} : \beta \otimes [D, x, H] \mapsto \beta \otimes [D, xg, H^g]\ ,$$

for all β in B_D, and $D^x \leqslant H$.

Proposition 5.61. (a) $\langle\mathbf{B}\rangle$ *is a G-functor into* $\mathfrak{M}_k$.

(b) *If* $\theta : \mathbf{B} \to \mathbf{B}'$ *is a morphism of G-semifunctors, then the maps*

$$\langle\theta\rangle_H : \Sigma\, \beta \otimes [D, x, H] \mapsto \Sigma\, \beta\theta_D \otimes [D, x, H]\ ,\quad (\beta \in B_D,\ D^x \leqslant H)\,,$$

define a morphism $\langle\theta\rangle : \langle\mathbf{B}\rangle \to \langle\mathbf{B}'\rangle$, *and the correspondence* $\mathbf{B} \to \langle\mathbf{B}\rangle$, $\theta \to \langle\theta\rangle$ *defines a functor from* $\mathfrak{M}_k^0(G) \to \mathfrak{M}_k(G)$.

(c) *If* $\mathbf{A} : \mathcal{S}(G) \to \mathfrak{M}_k$ *is a G-functor, then the maps*

$$\psi_H : \Sigma\, \alpha \otimes [D, x, H] \mapsto \Sigma\, \alpha T(D, x, H) \quad (\alpha \in A_D,\ D^x \leqslant H)$$

define a surjective morphism $\psi : \langle \mathbf{A}_0 \rangle \to \mathbf{A}$.

(d) *For any G-semifunctor* $\mathbf{B} : \mathcal{S}_0(G) \to \mathcal{M}_k$, $\langle \mathbf{B} \rangle$ *is a relative projective object in* $M_k(G)$, *relative to the class of "R-split epimorphisms" ; an R-split epimorphism is defined to be a surjective morphism* $\theta : \mathbf{X} \to \mathbf{Y}$ *of G-functors, such that there exists a morphism* $\tau : \mathbf{Y}_0 \to \mathbf{X}_0$ *of G-semifunctors for which* $\tau\theta = \mathrm{id}_{\mathbf{Y}}$.

(e) *(Adjoint property of* $\langle \mathbf{B} \rangle$, $\mathbf{A}_0$.*) For a given G-semifunctor* $\mathbf{B} : \mathcal{S}_0(G) \to \mathcal{M}_k$, *define the monomorphism* $\mu : \mathbf{B} \to \langle \mathbf{B} \rangle_0$ *of semifunctors by* $\mu_H : \beta \mapsto \beta \otimes [H, 1, H]$ $(H \leqslant G, \beta \in B_H)$. *Then if* $\mathbf{A}$ *is any G-functor* $\mathbf{A} : \mathcal{S}(G) \to \mathcal{M}_k$, *and if* $\theta : \mathbf{B} \to \mathbf{A}_0$ *is any morphism of G-semifunctors, there is a unique morphism* $\bar{\theta} : \langle \mathbf{B} \rangle \to \mathbf{A}$ *of G-functors such that* $\theta = \mu\bar{\theta}\ (\mathrm{id}_{\mathbf{A}})$.

The proof of this proposition is omitted. By taking for $\mathbf{B}$ suitable "free G-semifunctors", we can make "free G-functors" $\langle \mathbf{B} \rangle$. We leave this as an exercise for the reader.

References

[1] M.F. Atiyah, Characters and cohomology of finite groups, Inst. Hautes Etudes Sci. Publ. Math. 9 (1961) 23–64.

[2] M.F. Atiyah and I.G. Macdonald, Introduction to commutative algebra (Addison-Wesley, London, 1969).

[3] R. Brauer, Zur Darstellungstheorie der Gruppen endlicher Ordnung I, Math. Z. 63 (1956) 406–444.

[4] R. Brauer and J.Tate, On the characters of finite groups, Ann. of Math. 62 (1955) 1–7.

[5] H.Cartan and S. Eilenberg, Homological algebra (Princeton, 1956).

[6] C.W. Curtis and I. Reiner, Representation theory of finite groups and associative algebras (Wiley, New York, 1962).

[7] A. Dress, On relative Grothendieck rings, Bull. Am. Math. Soc. 75 (1969) 955–958.

[8] A. Dress, On integral representations, Bull. Am. Math. Soc. 75 (1969) 1031–1034.

[9] B. Eckmann, Cohomology of groups and transfer, Ann. of Math. 58 (1953) 481–493.

[10] J.A. Green, On the converse to a theorem of R. Brauer, Proc. Cambridge Philos. Soc. 51 (1955) 237–239.

[11] J.A. Green, On the indecomposable representations of a finite group, Math. Z. 70 (1959) 430–445.

[12] J.A. Green, Blocks of modular representations, Math. Z. 79 (1962) 100–115.

[13] J.A. Green, The modular representation algebra of a finite group, Illinois J. Math. 6 (1962) 607–619.

[14] J.A. Green, A transfer theorem for modular representations, J. Algebra 1 (1964) 73–84.

[15] J.A. Green, Some remarks on defect groups, Math. Z. 107 (1968) 133–150.

[16] N. Jacobson, Structure of rings (Providence, R.I., 1956).

[17] D.L. Johnson, A transfer theorem for the cohomology of a finite group, Invent. Math. 7 (1969) 174–182.

[18] T.Y. Lam, Induction theorems for Grothendieck groups and Whitehead groups of finite groups, Ann. Sci. Ecole Norm. Sup., 4e série, 1 (1968) 91–148.

[19] T.Y. Lam and I. Reiner, Relative Grothendieck groups, J. Algebra 11 (1969) 213–242.
[20] T.Y. Lam and I. Reiner, Reduction theorems for relative Grothendieck rings, Trans. Am. Math. Soc. (1970), to be published.
[21] T.Y. Lam and I. Reiner, Restriction maps on relative Grothendieck rings, J. Algebra (1970), to be published.
[22] P. Roquette, Arithmetische Untersuchung des Charakterringes einer endlichen Gruppe, J. Reine Angew. Math. 190 (1952) 148–168.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 1 (1971) pp. 79–95.

LOWER BOUNDS FOR COMPLEXITY OF FINITE SEMIGROUPS*

John RHODES ‡ and Bret R.TILSON
Department of Mathematics, University of California, Berkeley, Calif., USA
and Department of Mathematics, Columbia University, New York, N.Y., USA

Received 23 December 1969

We assume the reader is familiar with the definition and elementary properties of the (group) complexity of a finite semigroup or of a finite state sequential machine. See [1] chapters 6 and 9, or [6], or [7].

All semigroups considered are finite unless it is explicitly stated otherwise. We also assume the reader is acquainted with the terminology and results of [1] chapters 1, 5–9.

Herein, we derive the strongest lower bounds $\#_I(S)$ (defined below) achieved to date for the (group) complexity, $\#_G(S)$, of the finite semigroup S, $\#_I(S) \leq \#_G(S)$. In all examples and cases known to the authors $\#_I(S) = \#_G(S)$ and we conjecture (with some hope) $\#_I(S) = \#_G(S)$ for all finite semigroups S. The conjecture is known to be true if S is a union of groups (see chapter 9 of [1]) or more generally if S is regular and $S^{\mathcal{J}'}$ is combinatorial. Furthermore it is true if S has at most two non-zero $\mathcal{J}$ classes. $\#_I(S)$ is defined as follows. T is said to be a "T_1" *semigroup* ‡‡ iff T is generated by a chain $L_1 > L_2 > \ldots > L_n$ of certain of its $\mathcal{L}$-classes. Here as usual $L_1 > L_2$ iff $S^1 L_1 \supset S^1 L_2$. $\mathrm{EG}(T)$ is the subsemigroup of T generated by its idempotents. Then $\#_I(S)$ is, by definition, the largest non-negative integer n so that there exists a chain of subsemigroups

$$S \geq T_1 \geq \mathrm{EG}(T_1) \geq T_2 \geq \mathrm{EG}(T_2) \geq \ldots \geq T_n \geq \mathrm{EG}(T_n)$$

where T_j is a non-combinatorial "T_1" semigroup for $j = 1, \ldots, n$. *We then prove* $\#_I(S) \leq \#_G(S)$ *for all finite semigroups S.*

On the road to proving this inequality we also prove that *all inverse semigroups have (group) complexity* ≤ 1. We also show that if the irreducible semigroup

* This research was sponsored in part by the United States Air Force, Office of Scientific Research, Grant No. AF-AFOSR-68-1477.

‡ Alfred P.Sloan Research Fellow.

‡‡ A variation of type I semigroup introduced in Definition 9.2.4.(i) of [1]. In fact this paper is a generalization to arbitrary finite semigroups of $\theta = \theta_i$ proved in chapter 9 of [1].

$U_1 = \{a, b\}^r$ is not a subsemigroup of S (or as we will say S is an "R_1" semigroup since there is at most one idempotent in any $\mathcal{R}$ class), then again $\#_G(S) \leq 1$.

We say that the semigroup S *commutes over groups* iff for each group G, there exists a group G' and a combinatorial semigroup C' so that $G \le S$ divides $C' \le G'$. We also show that "R_1" *semigroups are exactly the semigroups which commute over groups. Thus inverse semigroups commute over groups* since they are "R_1" semigroups.

As an application to automata theory we conclude that any cascade combination of Abelian machines (machine equals finite state sequential machine) can be performed by a cascade of a single group machine followed by a single combinatorial machine. This follows from the above, since Abelian semigroups are "R_1" semigroups and "R_1" is closed under wreath products and division.

Finally, given a collection of machines $\mathcal{M}$ whose semigroups are *monoids* (i.e. have identities) then exactly one of the following occur:

(1) Every cascade combination of members of $\mathcal{M}$ has complexity zero, or equivalently, all the semigroups of members of $\mathcal{M}$ are combinatorial;

(2) Cascade combinations of members of $\mathcal{M}$ having arbitrary large complexity exist;

(3) Every cascade combination of members of $\mathcal{M}$ has complexity ≤ 1 and some member of $\mathcal{M}$ has complexity 1, or equivalently, the semigroup of each member of $\mathcal{M}$ is an "R_1" semigroup, and at least one is non-combinatorial.

In particular, if the complexity of cascade combinations of a collection of machines is bounded by k, then k is less than or equal to 1.

§ 1. Complexity of inverse semigroups

In the following all semigroups are assumed to have finite order.

We recall that an inverse semigroup S is a semigroup for which each element $s \in S$ has a unique semigroup inverse denoted s^{-1}. That is, $s = ss^{-1}s$ and $s^{-1} = s^{-1}ss^{-1}$. (See [2] chapter 1 or [1] chapter 7.) An inverse semigroup is regular, each $\mathcal{L}$ class has a unique idempotent, each $\mathcal{R}$ class has a unique idempotent, and the number of $\mathcal{R}$ classes equals the number of $\mathcal{L}$ classes in each $\mathcal{J}$ class. The reader can refer to [1] chapters 5–9 for an exposition of the definition and concepts employed next.

1.1. Proposition. Let S be an inverse semigroup of order n. Let S_n be the symmetric group on n letters, let $X_n = \{1, ..., n\}$, and let $U_2 = \{0\}^I$. Then

$$S \mid (U_2, U_2) \le w (X_n, S_n)$$

and therefore $C(S) \leq (2, \mathbf{G})$. In particular, $\#_G(S) \leq 1$.

Proof. It is well known (see [2] Theorem 1.20) that S is isomorphic to a subsemigroup of the symmetric inverse semigroup on n letters denoted $\mathrm{SIS}_R(X_n)$. $\mathrm{SIS}_R(X_n)$ is the set of all partial 1:1 maps on n letters (see either [2] chapter 1 or [1] chapter 1, Ex. 1.4j).

We define an onto homomorphism φ from $(U_2, U_2) \mathrm{w} (X_n, S_n) \cong F(X_n, U_2) \times_Y S_n$ into $\mathrm{SIS}_R(X_n)$ as follows: Let $f \in F(X_n, U_2)$, $g \in S_n$. Then $\varphi(f, g) = \bar{f} \in \mathrm{SIS}_R(X_n)$ where the domain of $\bar{f}$ is the set $\{i \in X_n : f(i) = I\}$ and

$$\bar{f}(i) = (i)g \quad \text{for all } i \text{ in the domain of } \bar{f}.$$

It is easy to verify that φ is an onto homomorphism. Thus $S \mid \mathrm{SIS}_R(X_n) \mid (U_2, U_2) \mathrm{w} (X_n, S_n)$ and $C(S) \leq (2, \mathbf{G})$.

The above proposition vindicates the intuition that "inverse semigroups are 'nearly' groups".

1.2. Notation. Let G be a group. Denote by $\mathcal{RCM}(n, G)$ the semigroup of $n \times n$ matrices with entries in G^0 that are both row monomial and column monomial. That is, matrices which have at most one non-zero entry in each row and each column. Clearly $\mathcal{RCM}(n, G)$ is an inverse semigroup since for $[a_{ij}] \in \mathcal{RCM}(n, G)$, $[a_{ji}^{-1}]$ is the unique semigroup inverse of $[a_{ij}]$, where $0^{-1} = 0$ by convention.

1.3. Proposition. Let S be an RM semigroup (see [1] chapter 8, Definition 2.14) whose distinguished ideal I is an inverse semigroup. Then S is an LM semigroup with respect to I. Furthermore, S is a subsemigroup of an inverse semigroup, so $C(S) \leq (2, \mathbf{G})$ and $\#_G(S) \leq 1$.

Proof. I is 0-simple and regular, so give I a Rees matrix representation $\mathcal{M}^0(G;A,B;C)$ where G is a maximal subgroup of $J = I - \{0\}$, the non-zero $\mathcal{J}$ class of S in I. Since by assumption I is inverse, $|A| = |B|$ and C is the identity matrix. Given this representation of I, we now can represent S faithfully as a semigroup of $n \times n$ row monomial matrices over G, where $n = |B|$ (see section 2 of chapter 8 of [1]). We will show that each element of S (as a matrix) is also column monomial.

Let $s \in S$ and let R_s be the row monomial matrix associated with s. Also, for each $s \in S$, there exists a column monomial matrix L_s over G that describes how s acts on I by left multiplication. This representation need not in general be faithful.

The matrix form of the "linked equation" (see Fact 2.14 and Problem X2.15 of chapter 7 of [1]) says that for each $s \in S$

$$R_s C = C L_s .$$

But C is the $n \times n$ identity matrix so $R_s = L_s$ and R_s is therefore both row and column monomial. Thus the representation $s \to L_s$ is faithful and S is an LM semigroup.

Thus S has a faithful representation as a subsemigroup of the inverse semigroup $\mathcal{R}\mathcal{C}\mathcal{M}(n, G)$ and thus $C(S) \leq (2, \mathbf{G})$ by Proposition 1.1.

1.4. Remark. (a) Notice $\mathcal{R}\mathcal{C}\mathcal{M}(n, \{1\}) = \mathrm{SIS}_R(X_n)$ and $\mathcal{R}\mathcal{C}\mathcal{M}(n, G)$ satisfies the hypothesis of Proposition 1.3 for all n and G, by taking I to be all those $n \times n$ matrices with at most one non-zero entry in the entire matrix.

(b) Let S satisfy the hypothesis of Proposition 1.3 and further suppose I is a non-combinatorial ideal (e.g. $\mathcal{R}\mathcal{C}\mathcal{M}(n,G)$, $n \geq 1$, $G \neq \{1\}$). Then (1.3) proves that S is a GM semigroup and $(\mathbf{G}, 1) \leq C(S) \leq (\mathbf{C}, 2)$; thus if S is not a group, we have $C(S) = (\mathbf{C} \vee \mathbf{G}, 2)$ or $(\mathbf{C}, 2)$. Contrast this with the fact that if S is a GM semigroup with respect to the ideal I, and S is not a group but I is a union of groups, then $C(S) = (\mathbf{G}, k)$ for some $k \geq 2$. (See chapter 9 of [1].)

We have shown that $\#_G$(inverse semigroups) ≤ 1. We will strengthen this in the next section by showing that even semidirect or wreath products of inverse semigroups still have complexity ≤ 1.

§2. "R_1" Semigroups

2.1. Definition and remarks. (a) An "R_1" *semigroup* is a semigroup in which no two distict idempotents are $\mathcal{R}$ equivalent, i.e. each $\mathcal{R}$ class contains at most one idempotent. For example, inverse semigroups are "R_1" semigroups. Also, all Abelian semigroups are "R_1" semigroups.

(b) By U_1 we denote the semigroup $\{a, b\}^r$. Clearly, S is an "R_1" semigroup iff $U_1 \not\subseteq S$. But in [1] chapter 5, in the proof of Lemma 3.6, it is shown that $U_1 \subseteq S$ iff $U_1 \mid S$, and that $U_1 \in \mathrm{IRR}$. That is to say, if $U_1 \mid (X_2, S_2) \le (X_1, S_1)$ then $U_1 \mid S_1$ or $U_1 \mid S_2$. It is then easily concluded that the property "R_1" is closed under wreath products and division.

Consider any regular $\mathcal{J}$ class J of an "R_1" semigroup S. If J^0 is given a Rees matrix representation, then the structure matrix C of J^0 can be normalized to zeros and 1's (1 being the identity of the structure group) since each column of C has at most one non-zero entry. (See [1] chapter 7.) Furthermore, if we apply the homomorphism RM_J (and hence GM_J and RLM_J) to S, the image of $J \cup F(J)$, i.e. the distinguished ideal of $\mathrm{RM}_J(S)$, will be an inverse semigroup. To see this, it is necessary to recall that RM_J (and hence GM_J and RLM_J) identifies columns of the structure matrix of J^0 that are proportional (on the right). (See [1] 8.2.22.) A little reflection will show that the only form (after perhaps permuting rows and columns) that a structure matrix for J^0 can take (since it is an "R_1" semigroup) is that shown in fig. 1, so that under RM_J the structure matrix must collapse down to an identity matrix.

$$\begin{array}{cl}
 & \qquad\qquad\qquad\qquad A \\
 & \quad \overbrace{}^{1} \quad \overbrace{}^{2} \quad \overbrace{}^{3} \quad\ \ \overbrace{}^{n} \\
B = \begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ n \end{array} & \begin{bmatrix}
1 \cdots 1 & 0 \cdots 0 & 0 \cdots 0 & \cdots & 0 \cdots 0 \\
0 \cdots 0 & 1 \cdots 1 & 0 \cdots 0 & \cdots & 0 \cdots 0 \\
0 \cdots 0 & 0 \cdots 0 & 1 \cdots 1 & \cdots & 0 \cdots 0 \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & & \vdots \quad \vdots \\
0 \cdots 0 & 0 \cdots 0 & 0 \cdots 0 & \cdots & 1 \cdots 1
\end{bmatrix}
\end{array}$$

Fig. 1. $B \times A$ structure matrix of J.

Thus we have proved the following (by Proposition 1.3).

2.2. Lemma. If an RM semigroup is an "R_1" semigroup, then it is a subsemigroup of an inverse semigroup.

Recall ([1] chapter 6) that if S is a subdirect product of semigroups $S_1, \ldots, S_n$ (written $S \leq\leq S_1 \times \ldots \times S_n$) then $C(S) = \mathrm{Lub}\,\{C(S_i): i = 1, \ldots, n\}$ and $\#_G(S) = \max\{\#_G(S_i): i = 1, \ldots, n\}$. We refer to this fact as *Axiom I for complexity.*

2.3. Theorem. (a) The class of "R_1" semigroups is closed under division, semidirect and wreath products.

(b) Let S be an "R_1" semigroup. Then $C(S) \leq (2, \mathbf{G})$.

Proof. The assertion of (a) was proved in 2.1(b). We now prove (b) by induction on $|S|$. Assume (b) is true for all semigroups whose order is less than n, and let $|S| = n$. If S is not subdirectly indecomposable, then $S \leq\leq S_1 \times \ldots \times S_n$ where $|S_i| < |S|$, $i = 1, \ldots, n$, so by induction and Axiom I for complexity, $C(S) \leq (2, \mathbf{G})$.

Thus we can assume S is subdirectly indecomposable. By Proposition 8.2.12 of [1], S is either a GM, RLM, or LLM semigroup or S has a unique 0-minimal ideal I that is null. If S is GM or RLM, then S is an RM semigroup, so by Lemma 2.2, $C(S) \leq (2, \mathbf{G})$.

The proof of the theorem for the remaining cases will be handled by a series of lemmas that utilize the machine methods of [1] chapter 5. We will construct the machine S^f out of other machines whose semigroups are of known complexity in such a way as to prove the theorem. The first lemma shows that it is enough to be able to construct a machine f that computes as S^f except at zero. More precisely, we have

2.4. Lemma. Let S be a semigroup with a zero, and suppose there exists a machine $f : \Sigma S \to S$ such that if $S^f(\alpha) \neq 0$ for $\alpha \in \Sigma S$, then $f(\alpha) = S^f(\alpha)$. Then

$$C(S) \leq (1, \mathbf{C}) \oplus C(f^S)\,.$$

Proof. Consider the machine equation

$$S^f = (S^{\#})^0\, k^{\Gamma}\, 2^{\sigma}_{S \times S}(S^{rf} \times f)^{\sigma} \Delta^{\Gamma} \tag{1}$$

where $S^{\#}$ is the set $S - \{0\}$, and

$$\Delta : S \to S \times S \quad \text{with } \Delta(s) = (s, s)$$

and

$$k : (S \cup \{*\}) \times S \to S$$

with

$$k[*, (s_1, t_1)] = s_1$$

and

$$k[(s_{n-1}, t_{n-1}), (s_n, t_n)] = t_{n-1} s_n\,.$$

We verify equation (1); for notation, let $\alpha_k = (s_1, \ldots, s_k)$, $k = 1, \ldots, n$, and let $t_k = f(\alpha_k)$. Consider the sequence α_n and suppose k, $1 \leq k \leq n$, is the smallest integer such that $S^f(\alpha_k) = 0$. Then $S^f(\alpha_i) = f(\alpha_i) = t_i$ for $i = 1, \ldots, k-1$, by assumption. Thus $t_{i-1} s_i = t_i = S^f(\alpha_i)$ for $i = 2, \ldots, k-1$, and therefore equation (1) is correct for inputs $\alpha_1, \ldots, \alpha_{k-1}$. Now, since $t_{k-1} = S^f(\alpha_{k-1})$, we have $t_{k-1} s_k = S^f(\alpha_{k-1}) s_k =$

$S^f(\alpha_k) = 0$. Then equation (1) gives 0 for input α_k and all inputs thereafter regardless of the values of $f(\alpha_i)$, $i \geq k$. Thus equation (1) is verified. Now using the relations between machine equations and wreath products of chapter 5 of [1] we have

$$S \mid C \le f^S$$

for some combinatorial semigroup C, since $(S^\#)^{r0}$, $2^S_{S\times S}$, and S^r are combinatorial. Thus

$$C(S) \leq (1, C) \oplus C(f^S) .$$

2.5. Remark. Let S be an "R_1" RM semigroup with distinguished ideal $I = \mathcal{M}^0(G;A,B;C)$. By Proposition 1.3 we can consider S as a semigroup of row monomial matrices $(n\times n)$ over G^0. Define a map $\psi : S \to G_n$, the group of units of $\mathcal{R}\mathcal{M}(n, G)$, (i.e. the elements of $\mathcal{R}\mathcal{M}(n, G)$ with exactly one entry in each row and column) as follows: Take the matrix s and extend it arbitrarily to a matrix of G_n, calling that element $\psi(s)$. That is, enter into each zero row of s a group entry in such a way as to obtain an element of G_n. This can be done since $s \in S$ is column monomial by Proposition 1.3.

Notice that $[(G \times B) \cup \{0\}, S]$ and $[(G \times B) \cup \{0\}, G_n]$ are faithful transformation semigroups.

2.6. Lemma. Given the above situation, let $(g, b) \in G \times B$ and $s, s_1, s_2 \in S$. Then

(a) If $(g, b)s \neq 0$, then $(g, b)s = (g, b)\psi(s)$.

(b) Let $(g, b)s_1 s_2 \neq 0$, then $(g, b)\psi(s_1 s_2) = (g, b)\psi(s_1)\psi(s_2) = (g, b)s_1 s_2$.

Proof. (a) follows from the way $\psi : S \to G_n$ was defined. For (b), if $(g, b)s_1 s_2 \neq 0$, then $(g, b)s_1 \neq 0$ so by (a) $(g, b)s_1 = (g, b)\psi(s_1)$. Now again by (a), $(g, b)\psi(s_1 s_2) = (g, b)s_1 s_2 = [(g, b)\psi(s_1)] s_2 = (g, b)\psi(s_1)\psi(s_2)$.

We now introduce the *right activator* of a $\mathcal{J}$ class, a construction by which global questions concerning a null $\mathcal{J}$ class of a semigroup can sometimes be reduced to the study of a regular $\mathcal{J}$ class.

2.7. Fact. Let J be a $\mathcal{J}$ class of a semigroup S. Let $\alpha(J) = \{s \in S : Js \cap J \neq \emptyset\}$. Then

(a) $\alpha(J)$ is a union of $\mathcal{J}$ classes of S and if J_1 and J_2 are $\mathcal{J}$ classes of S such that $J_1 \subseteq \alpha(J)$ and $J_1 \leq J_2$, then $J_2 \subseteq \alpha(J)$.

(b) $\alpha(J)$ has a unique $(\leq)$ minimal $\mathcal{J}$ class which is regular.

Proof. To prove (a), we note that the complement of $\alpha(J)$ in S, that is $\{s \in S : Js \cap J = \emptyset\}$, is an ideal. Thus, the complement, and hence $\alpha(J)$ are unions of $\mathcal{J}$ classes. The second statement follows easily, since $\alpha(J)$ is the complement of an ideal.

(b) Let J_1 be a $(\leq)$ minimal $\mathcal{J}$ class of $\alpha(J)$. We show first that for each $j \in J$, there

exists $s \in J_1$ such that $js \in J$. Let $s \in J_1$. Then there exists $j \in J$ such that $js \in J$. If $j' \mathcal{L} j$, then there exists $x \in S$ such that $j = xj'$. Now $js = xj's \in J$ so $j's \in J$. Suppose $j' \mathcal{R} j$. Then there exists $y \in S$ such that $j = j'y$, and $js = j'ys \in J$. Then $ys \in \alpha(J)$, and so $ys \in J_1$ by minimality and $j'(ys) \in J$. Thus every element of J is "kept up" by multiplication by some element of J_1.

Let $j \in J$ and $s \in J_1$ be such that $js \in J$. Then by the above there exists $t \in J_1$ such that $jst \in J$ which implies $st \in J_1$ by minimality. Thus J_1 is regular.

Finally, suppose J_1 and J_2 are two distinct minimal $\mathcal{J}$ classes of $\alpha(J)$. Then $J_1J_2 \cap \alpha(J) = \emptyset$. Let $j \in J$ and $s_1 \in J_1$ be such that $js_1 \in J$. Since J_2 is minimal, there exists $s_2 \in J_2$ such that $js_1s_2 \in J$, so $s_1s_2 \in \alpha(J)$, a contradiction. Thus $\alpha(J)$ has a unique minimal $\mathcal{J}$ class that is regular.

2.8. Definition. The *right activator of* J, $\mathrm{RA}(J)$, is the unique ($\leq$) minimal $\mathcal{J}$ class of S contained in $\alpha(J)$. By Fact 2.7, $\mathrm{RA}(J)$ is well defined and regular. Of course, $\mathrm{RA}(J) = J$ iff J is regular.

2.9. Lemma. Let J be a $\mathcal{J}$ class of a semigroup S and let $J_1 = \mathrm{RA}(J)$. Then

(a) For each $j \in J$, there exists an (idempotent) element $e \in J_1$ such that $j = je$.

(b) Define the (not necessarily faithful) transformation semigroups (J^0, S) and (J_1^0, S) in the usual way. For each $j \in J$, choose an element $e_j \in J_1$ such that $j = je_j$. Make (J_1^0, S) faithful by applying RM_{J_1} to S. Then for all $j \in J$ and $s \in S$ we have

$$j \cdot s = j(e_j \cdot \mathrm{RM}_{J_1}(s))$$

where if $e_j \cdot \mathrm{RM}_{J_1}(s) = 0 \in J_1^0$, it is understood that $j(e_j \cdot \mathrm{RM}_{J_1}(s)) = 0 \in J^0$.

Proof. (a) Let $j \in J$. Then there exists $s \in J_1$ such that $js \in J$. Then $js \mathcal{R} j$, so there exists $t \in S$ such that $(js)t = j$. By minimality of J_1, $st \in J_1$. Let $e = st$, or if an idempotent is desired, note that

$$j = j(st) = j(st)^2 = \ldots = j(st)^n ,$$

so there exists an integer n such that $(st)^n$ is an idempotent belonging to J_1.

(b) If $e_j \cdot \mathrm{RM}_{J_1}(s) = 0 \in J_1^0$, then $e_js \in S - \alpha(J)$, by the minimality of J_1 in $\alpha(J)$. Hence $js = j(e_js) = 0$. If $e_j \cdot \mathrm{RM}_{J_1}(s) \in J_1$, then $j(e_j \cdot \mathrm{RM}_{J_1}(s)) = j(e_js) = js$.

2.10. Remark. Let $j \in J$ and let e be an idempotent in J_1 such that $j = je$. Let $R[x]$ denote the $\mathcal{R}$ class containing x. Then for each $x \in S$, $(R[x]^0, S)$ is a (not necessarily faithful) transformation semigroup, where the action is defined in the usual way. Then

$$(R[e]^0, S) \overset{(\theta, \mathrm{Id})}{\twoheadrightarrow} (R[j]^0, S)$$

where Id : $S \twoheadrightarrow S$ is the identity map. $\theta : R[e]^0 \twoheadrightarrow R[j]^0$ is defined by $\theta(s) = js$ for $s \in R[e]$ and $\theta(0) = 0$. Furthermore $\theta(x) = 0$ iff $x = 0$.

Proof. Let $x \in R[e]$. Then $x \mathcal{R}\ e$, so $jx \mathcal{R}\ je = j$. In other words $jx = \theta(x) \in R[j]$. Therefore θ is well-defined and $\theta(x) = 0$ iff $x = 0$. To show θ is onto, let $j' \in R[j]$. Then there exists $s \in S^1$ such that $j' = js$, so $j' = js = jes$ and $es \in J_1$ by minimality. But $es \mathcal{R}\ e$, so $\theta(es) = jes = j'$ and θ is onto. We must verify that for all $x \in R[e]^0$ and for all $s \in S$, $\theta(x)s = \theta(xs)$. If $x = 0$, then $xs = 0$, so $\theta(x)s = \theta(xs)$ in that case. If $x \in R[e]$ and $xs = 0$, then $xs \in S - \alpha(J)$, i.e. xs annihilates J. Thus $\theta(x)s = (jx)s = j(xs) = 0 = \theta(xs)$. If $xs \in R[e]$, then $x \in R[e]$ and $\theta(xs) = jsx = (jx)s = \theta(x)s$.

Part (b) of Lemma 2.9 is a particular application of the above.

We are now ready to finish the proof of the theorem. The critical idea is that the action of S on J (the minimal J class in the two remaining cases) can be computed by the action of $\mathrm{RM}_{J_1}(S)$ on $J_1 = \mathrm{RA}(J)$, and since S is an "R_1" semigroup, the action of $\mathrm{RM}_{J_1}(S)$ on J_1 can be computed (up to zero) by a *group*, by Lemma 2.6. Then we are in position to finish by using Lemma 2.4.

Proof of Theorem 2.3. The cases that remain are (1) S is an LLM semigroup with distinguished ideal I and (2) S has a unique 0-minimal null ideal I. In both cases, let $J = I - \{0\}$ be the non-zero J class of S contained in I. Let $J_1 = \mathrm{RA}(J)$. $\mathrm{RM}_{J_1}(S)$ is an "R_1" RM semigroup, so let $\psi : \mathrm{RM}_J(S) \to G$, a group, be defined as in Remark 2.5. Let $\theta = \psi \cdot \mathrm{RM}_{J_1}$ and let $\eta : S \twoheadrightarrow S/I$ be the natural epimorphism associated with I. Finally, for each $j \in J$ let $e_j \in J_1$ be chosen so that $je_j = j$, as in Lemma 2.9(b).

Now define the machine $f : \Sigma S \to S$ by

$$f = h_3(S^{rIf} \times G^{rIf} \times G^{rf})\, h_2^{\Gamma} 2_X^{\sigma}(S^{rf} \times G^f \times S/I^f)^{\sigma} h_1^{\Gamma} \tag{2}$$

where $X = S \times G \times S/I$ and

1. $h_1 : S \to X$ with $h_1(s) = (s, \theta(s), \eta(s))$.
2. $h_2 : (X \cup \{*\}) \times X \to S^I \times G^I \times G$ with

$$h_2[*, (s, g, t)] = \begin{cases} (t, I, g) \text{ if } t \neq 0\,, \\ (s, g, g) \text{ if } t = 0\,, \end{cases}$$

$$h_2(s_1, g_1, t_1), (s_2, g_2, t_2)] = \begin{cases} (t_2, I, g_2) & \text{if } t_2 \neq 0 \\ (t_1 s_2, g_2, g_2) & \text{if } t_2 = 0 \text{ and } t_1 \neq 0 \\ (I, I, g_2) & \text{if } t_2 = 0 \text{ and } t_1 = 0. \end{cases}$$

3. $h_3 : S^I \times G^I \times G \to S$ with

$$h_3(s,x,g) = \begin{cases} s & \text{if } x = I \\ 0 & \text{if } x \neq I \text{ and } s = 0 \\ s(e_s \cdot x^{-1} g) & \text{if } x \neq I \text{ and } s \neq 0 \text{ (in this case } s \in J). \end{cases}$$

Now, by induction, $C(S/I) \leq (2, \mathbf{G})$, so by equation (2) and chapter 5 of [1], $C(f^S) \leq (2, \mathbf{G})$.

But we claim that the machine f has the property that if $S^f(\alpha) \neq 0$, then $f(\alpha) = S^f(\alpha)$. Then by Lemma 2.4

$$C(S) \leq (1, \mathbf{C}) \oplus C(f^S) = (2, \mathbf{G}).$$

While this claim can be proved by induction on the length of a sequence of ΣS, we believe it is more instructive to give a hueristic discussion of why the machine f works, leaving the verification of the claim to the reader.

Let $\alpha_k = (s_1, ..., s_k) \in \Sigma S$, $k = 1, 2, ...$. Putting a sequence α_n into f, we see that as long as the product $s_1 \cdots s_k = S^f(\alpha_k)$ does not fall into I, the machine S/I^f will give the correct answer. Suppose $s_1 \cdots s_n \in I$ but $s_1 \cdots s_{n-1} \notin I$. Then the delay machine, $h_2^{\Gamma} 2_X^{\sigma}$, computes the actual value of $s_1 \cdots s_n$ in I, and this value is saved forever by the machine S^{rIf}. At the same time, the output, g, of the machine G^f is saved by the G^{rIf} machine. Since $s_1 \cdots s_n = s \in I$, the outputs of G^f at time $n+1$ and after determine what happens to s under multiplication by s_{n+1} if $ss_{n+1} \neq 0$. That is, $s[e_s \cdot g^{-1}(gg_{n+1})] = ss_{n+1} = s_1 \cdots s_{n+1}$ if $s_1 \cdots s_{n+1} \neq 0$. From this discussion of f, it is not hard to verify that if $S^f(\alpha) \neq 0$, then $f(\alpha) = S^f(\alpha)$ for all $\alpha \in \Sigma S$. This proves Theorem 2.3.

We can now state some interesting and important corollaries of Theorem 2.3. We first need some definitions.

2.11. Definition. (a) *S commutes over groups, denoted $S \in \mathcal{C}\,\mathcal{G}$* iff for all groups G,

$$G \le S \mid C' \le G'$$

for some combinatorial semigroup C' and some group G'.

(b) Let $\mathcal{S}$ denote a collection, possibly infinite, of finite semigroups. Then $\#_G(\mathcal{S}) = \max\{\#_G(S) : S \in \mathcal{S}\} \in [0, +\infty]$. We say $\mathcal{S}$ is bounded iff $\#_G(\mathcal{S}) = n < +\infty$. $W(\mathcal{S})$ denotes the wreath product and divisor closure of $\mathcal{S}$. See [1] 5.2.17.

2.12. Corollary. (a) $\mathcal{C}\,\mathcal{G}$ equals the "R_1" semigroups.

(b) Let $\mathcal{S}$ be a collection of *monoids* which are not all combinatorial semigroups, so $\#_G(\mathcal{S}) \neq 0$. Then the following statements are equivalent.

(i) $\#_G(W(\mathcal{S})) = 1$.

(ii) $\#_G(W(\mathcal{S}))$ is bounded.

(iii) $\mathcal{S} \subseteq \mathcal{C}\,\mathcal{G}$.

(iv) Each member of $\mathcal{S}$ is an "R_1" semigroup.

(v) $U_3 = \{a, b\}^{r1}$ is not a subsemigroup of any member of $\mathcal{S}$.

Proof. (a) Let S be an "R_1" semigroup, and consider $G \le S$. Since all groups are "R_1", $G \le S$ is also an "R_1" semigroup. Thus by Theorem 2.3 $C(G \le S) \leq (2, G)$, so $G \le S | C' \le G'$. Thus each "R_1" semigroup commutes over groups. Conversely if $U_1 | S$ then $G \le U_1 | G \le S$. But by [1] Theorem 6.2.10(a), $C(G \le U_1) = (2, C)$, and thus $G \le U_1$ never divides $C' \le G'$. Thus $U_1 | S$ implies $S \notin \mathcal{C}\,\mathcal{G}$. This proves (a).

(b) Let S be a *monoid*. Then we first notice that $\{a, b\}^r = U_1 | S$ iff $\{a, b\}^{r1} = U_3 | S$. Further $U_i \leq S$ iff $U_i | S$, $i = 1, 2, 3$, by [1], in the proof of 5.3.6. Thus (v) $\Longleftrightarrow$ (iv). By (a) (iii) $\Longleftrightarrow$ (iv). Theorem 2.3 gives us (iv) $\Rightarrow$ (i), and trivially (i) $\Rightarrow$ (ii). Thus we must show (ii) $\Rightarrow$ (v). Thus suppose $U_3 | S$ for some $S \in \mathcal{S}$. Then we must show $\#_G(W(\mathcal{S}))$ is unbounded. By assumption there exists for some $S_1 \in \mathcal{S}$ a non-trivial group G such that $G | S_1$. Thus (by chapter 5 of [1]), $T_n = G \le U_3 \le G \le U_3 \le \ldots \le G \le U_3$ (length $2n$) belongs to $W(\mathcal{S})$, and by Theorem 6.2.10(b) of [1], $\#_G(T_n) = n$. Thus $\#_G(W(\mathcal{S}))$ is unbounded.

2.13. Remark. (a) The restriction to monoids in the second part of the previous corollary was necessary, for there is a collection of semigroups closed under wreath products whose $\#_G$ is bounded by 1 that contain non-"R_1" semigroups. For instance, $W(\mathcal{G} \cup U_1)$, the wreath product-divisor closure of all groups $\mathcal{G}$ and $U_1 = \{a, b\}^r$. The semigroup $U_2 = \{0\}^l$ does not divide any member of $W(\mathcal{G} \cup U_1)$ because $U_2 \in \mathrm{IRR}$, and it is easy to see that any semigroup not divisible by U_2 is a nilpotent extension of a simple semigroup, or equivalently, a semigroup whose only regular $\mathcal{J}$ class is its kernel. These semigroups are easily shown to have $\#_G \leq 1$. For a proof, see [3] section 7.

Thus, we have $\#_G$ of $W(\mathcal{G} \cup U_1)$ bounded by 1 and $\#_G$ of $W(\mathcal{G} \cup U_2)$ bounded by 1 (since every member of $W(\mathcal{G} \cup U_2)$ is an "R_1" semigroup), and of course we know that $\#_G$ of $W(\mathcal{G} \cup U_3)$ is unbounded and, in fact, consists of all finite semigroups. We imagine that $\#_G$ of $W(\mathcal{G} \cup U_1 \cup U_2)$ is bounded by 1, i.e., we conjecture that $\#_G$ of the collection of semigroups not divisible by U_3 (closed under wreath products) is bounded by 1.

(b) Let S be an "R_1" semigroup. Then $\mathrm{EG}(S)$, the subsemigroup generated by the idempotents of S, is combinatorial.

Proof. By Theorem 2.3 $S | C' \le G'$. Then $\mathrm{EG}(S) | \mathrm{EG}(C' \le G')$ (using Fact 1.1.12 of [1]). But since $\mathrm{EG}(G') = \{1\}$, $\mathrm{EG}(C' \le G') \subseteq F(G', C')$, a combinatorial semigroup (see Lemma 6.2.7 of [1]). Thus $\mathrm{EG}(S)$ is combinatorial.

A direct proof, independent of Theorem 2.3 is also possible.

(c) From Lemma 2.2 we have that if S is an "R_1" semigroup, then $S \twoheadrightarrow S^{\mathrm{GM}}$ with S^{GM} a subdirect product of subsemigroups of inverse semigroups. Clearly $S \underset{\gamma}{\twoheadrightarrow} S^{\mathrm{GM}}$, and by Lemma 1.3 and Axiom I for complexity, $\#_G(S^{\mathrm{GM}}) \leq 1$.

Now if the "Fundamental Lemma of Complexity" were true (see chapter 9 of [1]) it would imply that $S \underset{\gamma}{\twoheadrightarrow} T$ implies $\#_G(S) = \#_G(T)$ (see [1] 9.3.4) which would then immediately imply (by (1.3) and (2.2)) that $\#_G$("R_1" semigroups) ≤ 1. Essentially (2.3) amounts to proving the Fundamental Lemma of Complexity for "R_1" semigroups, which we do with the aid of the right activator.

Currently Rhodes is writing up and checking a proof of the Fundamental Lemma. The proof uses the Zeiger construction of [4], appropriately modified, and the classification of epimorphisms given in [5].

And in terms of finite automata theory we have

2.14. Corollary ("Without flip-flops little can be done"). Let M be an arbitrary cascade composition of machines whose semigroups do not contain $\{a, b\}^r$ (for a good example, say Abelian). Then M can be replaced by a cascade of a group machine followed by a combinatorial machine.

Proof. Use the previous results and see chapters 3 and 6 of [1] for details.

§3. Lower bounds for group-complexity

In this section we show the existence of a function $\#_I$ from all finite semigroups, $\mathcal{S}$, into the non-negative integers, N, that is a lower bound for $\#_G$. That is, $\#_I(S) \leq \#_G(S)$ for all $S \in \mathcal{S}$. For every type of semigroup for which we can compute $\#_G$ to date, we find that $\#_I = \#_G$ so we hope to be able eventually to prove $\#_I = \#_G$ in general. We have already shown that $\#_I = \#_G(S)$ for all S which are union of groups. See (3.7)(b).

3.1. Definition. *S is a "T_1" semigroup* * iff T is generated by a chain $L_1 > L_2 > \ldots > L_n$ of its $\mathcal{L}$-classes. Here as usual $L_1 > L_2$ iff $S^1L_1 \supset S^1L_2$.

We observe that $F_R(X_n)$ is a "T_1" semigroup since it is generated by its group of units together with any idempotent of rank $n-1$.

3.2. Definition. Let S be a finite semigroup and consider chains of subsemigroups of S, $[S_1, \mathrm{EG}(S_1), S_2, \mathrm{EG}(S_2), \ldots, S_n, \mathrm{EG}(S_n)]$ where S_1 is a non-combinatorial "T_1" semigroup contained in S, S_2 is a non-combinatorial "T_1" semigroup contained in $\mathrm{EG}(S_1)$, ..., S_n is a non-combinatorial "T_1" semigroup contained in $\mathrm{EG}(S_{n-1})$. Let n be the length of such a series.

Define a function $\#_I : \mathcal{S} \to N$ by $\#_I(S)$ = maximum of the lengths of all such chains of subsemigroups of S.

3.3. Lemma. Let $\varphi : S \twoheadrightarrow T$ where T is a "T_1" semigroup generated by $L_1 > \ldots > L_n$. Then there exists a subsemigroup S' of S that is "T_1" and $\varphi(S') = T$. Conversely, if S is a "T_1" semigroup, then T is a "T_1" semigroup.

Proof. Let S_1 be a subsemigroup of smallest order subject to the condition $\varphi(S_1) = T$. Consider φ as the restriction of φ to S_1 from this point on.

Choose L_1^* to be an $\mathcal{L}$ class of S_1 such that $\varphi(L_1^*) = L_1$; L_1^* exists by Fact 7.2.1 of [1]. Now suppose $\mathcal{L}$-classes $L_1^*, \ldots, L_j^*$ of S_1 have been chosen such that $1 \leq j \leq n$, $\varphi(L_k^*) = L_k$, $k = 1, \ldots, j$ and $L_1^* > \ldots > L_j^*$. Then consider $\mathcal{L}_{j+1} = \{L'_{j+1} : L'_{j+1}$ is an $\mathcal{L}$ class of S_1, $\varphi(L'_{j+1}) \subseteq L_{j+1}$ and $L_j^* > L'_{j+1}\}$. We first show that $\mathcal{L}_{j+1}$ is non-empty.

Let $t \in T$ be chosen so that $tL_j \cap L_{j+1} \neq \emptyset$. Let $t' \in S_1$ such that $\varphi(t') = t$. Then $\varphi(t'L_j^*) = \varphi(t')\varphi(L_j^*) = tL_j$ and $tL_j \cap L_{j+1} \neq \emptyset$. Thus $\varphi^{-1}(L_{j+1}) \cap t'L_j^* \neq \emptyset$. Now $\varphi^{-1}(L_{j+1})$ is a union of $\mathcal{L}$ classes of S_1 and any $\mathcal{L}$ class of $\varphi^{-1}(L_{j+1})$ that meets $t'L_j^*$ belongs to $\mathcal{L}_{j+1}$. Thus $\mathcal{L}_{j+1}$ is non-empty.

Let L_{j+1}^* be a $(\leq)$ minimal member of $\mathcal{L}_{j+1}$. Then $L_j^* > L_{j+1}^*$ and $\varphi(L_{j+1}^*) \subseteq L_{j+1}$. But $S_1^1L_{j+1}^*$ is a left ideal, so $\varphi(S_1^1L_{j+1}^*) = T^1\varphi(L_{j+1}^*) = T^1L_{j+1}$, but by minimality of L_{j+1}^*, $\varphi(S_1^1L_{j+1}^* - L_{j+1}^*) \cap L_{j+1} = \emptyset$. Thus $\varphi(L_{j+1}^*) = L_{j+1}$.

Thus we can choose $L_1^* > \ldots > L_n^*$ so that $\varphi(L_k^*) = L_k$. But then if S_2 is the sub-

* A variation of type-I of chapter 9 of [1].

semigroup generated by $L_1^* \cup ... \cup L_n^*$ $(S_2 = \langle L_1^* \cup ... \cup L_n^* \rangle)$, we have $\varphi(S_2) = \langle \varphi(L_1^*) \cup ... \cup \varphi(L_n^*) \rangle = \langle L_1 \cup ... \cup L_n \rangle = T$. Thus by minimality of S_1, $S_2 = S_1$ so S_1 is a "T_1" semigroup. Thus take $S' = S_1$. The converse is clear.

3.4. Lemma. Let S be a combinatorial "T_1" semigroup. Then S is an "R_1" semigroup (and hence S commutes over groups.)

Proof. (An application of the "Principle of Induction for Combinatorial Semigroups". See 5.5 of [1].) Since S is combinatorial, $S \mid U_3^{(n)}$ for some $n \geq 1$. $(U_3^{(n)} = U_3 \mathrm{w} ... \mathrm{w}\, U_3$ (n times).) Assume $S \overset{\varphi}{\twoheadleftarrow} T \subseteq U_3^{(n)}$, where T is a "T_1" semigroup guaranteed by Lemma 3.3. It is sufficient to prove that T is an "R_1" semigroup. Induct on n. If $T \subseteq U_3$, then $T = \{0\}$ or $\{0\}^I$, the only "T_1" semigroups that are subsemigroups of U_3. Thus is this case T is an "R_1" semigroup.

Assume the lemma is true for n, that is, if $T \subseteq U_3^{(n)}$ then T is "R_1". Suppose $T \subseteq U_3^{(n+1)} \cong U_3^{(n)} \mathrm{w}\, U_3 \cong F(U_3, U_3^{(n)}) \times_Y U_3$, and suppose (f_0, r_0) and (f_1, r_1) are R equivalent idempotents of T. Let $p_1 : F(U_3, U_3^{(n)}) \times_Y U_3 \twoheadrightarrow U_3$ be the projection homomorphism. Then $p_1(T) \subseteq U_3$ is a "T_1" semigroup, so $p_1(f_0, r_0) = p_1(f_1, r_1)$, i.e. $r_0 = r_1$. $p_1(T)$ can either be one point or can be isomorphic to $\{0\}^I$, i.e., $p_1(T) = \{1, a\} \subseteq \{a, b\}^{r1}$, for example. Consider first the case when $p_1(T)$ is one point, say a. Then every element of T is of the form (f, a), where $f \in F(U_3, U_3^{(n)})$. So we can consider the homomorphism $\psi : T \to F(U_3, U_3^{(n)})$ defined by $\psi(f, a) = Y(a)(f) \equiv {}^a f$. Then $\psi(T) \subseteq F(U_3, U_3^{(n)})$ is "T_1" and hence is "R_1" by induction and the fact that $F(U_3, U_3^{(n)}) \cong U_3^{(n)} \times U_3^{(n)} \times U_3^{(n)}$. Now ${}^a f_0 = \psi(f_0, a) \mathcal{R}\ \psi(f_1, a) = {}^a f_1$ since ψ is a homomorphism, and thus ${}^a f_0 = {}^a f_1$ since $\psi(T)$ is "R_1". Now

$$(f_0, a) = (f_1, a)(f_0, a) = (f_1\, {}^a f_0, a) = (f_1\, {}^a f_1, a) = (f_1, a)^2 = (f_1, a),$$

i.e., $(f_0, a) = (f_1, a)$, so in this case T is "R_1".

Secondly, consider the case when $p_1(T) = \{0\}^I$ and $p_1(f_0, r_0) = p_1(f_1, r_1) = 0$. Then each element of T has for a first coordinate either 0 or I. The map $\psi : T \to F(U_3, U_3^{(n)})$ defined by $\psi(f, x) = Y(0) f$ is again a homomorphism. Hence the previous argument can be used again, proving that T is "R_1".

Finally, suppose $p_1(T) = \{0\}^I$ and $p(f_0, r_0) = p_1(f_1, r_1) = I$. Again each element of T has first coordinate either 0 or I. Let $L_1, ..., L_n$ be the $\mathcal{L}$ classes of T that generate T, and let $L_1, ..., L_k$ be the $\mathcal{L}$ classes whose elements have I in the first coordinate. Also, let $T' = \{(f, x) \in T : x = I\}$, a subsemigroup. Obviously (but critically) $T' = \langle L_1, ..., L_k \rangle$. In fact, it is also obvious (but critical) that each L_i, $i = 1, ..., k$ is an L class of T', so T' is a "T_1" semigroup. Now $p_1(T')$ is one element, so proceed as in the first case to conclude that T and hence S is an "R_1" semigroup.

3.5. Lemma. Let S be a non-combinatorial "T_1" semigroup. Then

(a) $\quad C(S) = (n, \mathbf{G})$ for some $n \geq 1$ or $C(S) = (2, \mathbf{C} \vee \mathbf{G})$

(b) $\quad \#_G(S) \geq 1 + \#_G[\mathrm{EG}(S)]$. $\qquad$ (3.1)

Proof. (a) Suppose $C(S) = (n, \mathbf{C})$, $n \geq 2$ or $(n, \mathbf{C} \vee \mathbf{G})$, $n \geq 3$. Then there exists (X_2, T) and (X_1, C) such that $S \twoheadleftarrow S_1 \subseteq (X_2, T) \le (X_1, C)$ where $C(T) = (n-1, \mathbf{G})$, C is a combinatorial semigroup, and S_1 is a "T_1" semigroup (by Lemma 3.3). Furthermore, let $p_1 : (X_2, T) \le (X_1, C) \twoheadrightarrow C$ be the projection homomorphism. Then we can assume $p_1(S_1) = C$, for if not, i.e., if $p_1(S_1) \subset C$, then $S_1 \subseteq (X_2, T) \le (X_1, p_1(S_1))$. But $p_1(S_1) = C$ is a combinatorial "T_1" semigroup, hence an "R_1" semigroup by the preceding lemma.

Now there are two cases. Let $n \geq 3$. Then write $(X_2, T) \le (X_1, C)$ as

$$(Y_2, T') \le (Y_1, G) \le (X_1, C)$$

where $C(T') = (n-2, \mathbf{C})$ and (Y_1, G) is a transformation group. But since C is "R_1", $(Y_1, G) \le (X_1, C)$ is "R_1", and hence has complexity $(2, \mathbf{G})$. Thus $C(S) \leq (n-2, \mathbf{C}) \oplus (2, \mathbf{G}) = (n-1, \mathbf{G})$, a contradiction.

Secondly, if $C(S) = (2, \mathbf{C})$, then

$$S \mid (X_2, G) \le (X_1, C).$$

But by the previous argument,

$$C[(X_2, G) \le (X_1, C)] = (2, \mathbf{G}).$$

Therefore $C(S) \leq (2, \mathbf{G})$, a contradiction. Hence (a) is proved.

(b) Recall that if $\mathrm{EG}(S) \neq \{0\}$, $C[\mathrm{EG}(S)] = (n, \mathbf{C})$ for some $n \geq 1$. (Use 6.2.7 of [1].) Let S be a non-combinatorial "T_1" semigroup. Assume $C(S) = (n, \mathbf{G})$, $n \geq 1$ and $C[\mathrm{EG}(S)] = (m, \mathbf{C})$, $m \geq 1$. Since

$$C[\mathrm{EG}(S)] \leq C(S),$$

we have $m + 1 \leq n$. In all cases we find that

$$\#_G(S) \geq 1 + \#_G[\mathrm{EG}(S)]. \qquad (3.1)$$

If $C(S) = (2, \mathbf{C} \vee \mathbf{G})$, then $C[\mathrm{EG}(S)] \leq (1, \mathbf{C})$. So again eq. (3.1) holds.

We are now ready to state and prove the theorem of this section, that $\#_I$ (Def. 3.2) is always a lower bound for $\#_G$.

3.6. Theorem. $\#_I(S) \leq \#_G(S)$ for all finite semigroups S.

Proof. If S is combinatorial, i.e. $\#_G(S) = 0$, then S has no "T_1" subsemigroups that

are non-combinatorial. Therefore, $\#_I(S) = 0$. Let S be non-combinatorial and assume $\#_I(S) = n$, where $n \geq 1$. Let the chain of subsemigroups

$$[S_1, \mathrm{EG}(S_1), ..., \mathrm{EG}(S_{n-1}), S_n, \mathrm{EG}(S_n)]$$

be a chain that achieves $\#_I$. Now by Lemma 3.5(b), $\#_G(S_i) \geq 1 + \#_G[\mathrm{EG}(S_i)]$, $i = 1, ..., n$, and since $S_{i+1} \subset \mathrm{EG}(S_i)$, $\#_G(S_{i+1}) \leq \#_G[\mathrm{EG}(S_i)]$. Then

$$\#_G(S_i) \geq 1 + \#_G(S_{i+1}), \quad i = 1, ..., n-1.$$

Therefore

$$\#_G(S_1) \geq 1 + \#_G(S_2) \geq 1 + 1 + \#_G(S_3) \geq ...$$

so

$$\#_G(S_1) \geq (n-1) + \#_G(S_n).$$

But $\#_G(S_n) \geq 1$, since S_n is non-combinatorial. Hence

$$\#_G(S) \geq \#_G(S_1) \geq n = \#_I(S).$$

3.7. Discussion and remarks. We conjecture that $\#_I(S) = \#_G(S)$ for all finite semigroups. Some evidence to date is the following:

(a) $\quad \#_I(F_R(X_n)) = \#_G(F_R(X_n)) = n-1$

and

$$\#_I(F_L(X_n)) = \#_G(F_L(X_n)) = n-1.$$

To see this let $S_n = F_R(X_n)$ or $F_L(X_n)$. Then S_n is a "T_1" semigroup since S_n is generated by its group of units G_1 together with any idempotent e_{n-1} of rank $n-1$. Further $\mathrm{EG}(S_n) = (S_n - G_1) \cup \{1\}$, so $S_{n-1} \subseteq \mathrm{EG}(S_n)$. Thus by induction, $n-1 \leq \#_I(S_n) \leq \#_G(S_n)$. But by Zeiger (see [8] section 5) $\#_G(F_R(X_n)) \leq n-1$ and by Allen [9], $\#_G(F_L(X_n)) \leq n-1$.

For $F_R(X_n)$, that $\#_I(F_R(X_n)) = \#_G(F_R(X_n))$, was first proved in [8] section 5.

We caution the reader that unlike the semigroup $F_R(X_n)$, $\#_G(S)$ and $\#_G(r(S))$ can differ by arbitrary amounts, even when S is a union of groups as examples of Zalcstein and Allen show. See [9] and [11].

(b) $\#_I(S) = \#_G(S)$ for all S which are unions of groups.

Proof. Essentially the proof of (i) of Theorem 9.2.5 of [1] yields (b). Specifically, in the proof of Lemma 9.2.29 we can clearly choose $y_n, ..., y_1$ so that $y_k^2 = y_k$ and $y_n > ... > y_1$. Let G_j be the maximal subgroup containing y_j and let J_j be the $\mathcal{J}$ class containing G_j. Then

$$X_n = G_n$$
$$X_{n-1} = J_{n-1}G_{n-1}G_n$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$X_1 = J_1G_1 \cdots G_n$$

And $T = X_n \cup \ldots \cup X_1$. But T is generated by $L[y_n] = G_n > L[y_{n-1}] = J_{n-1}G_{n-1} = J_{n-1}y_{n-1} > \ldots > L[y_1] = J_1G_1 = J_1y_1$. Thus T is a "T_1" semigroup, since each $L[y_k]$ is left simple (and thus is contained in an $\mathcal{L}$ class of T). Now,(b) follows from the proof in [1].

(c) Using the result mentioned in (2.13)(d) the authors can show that $\#_I(S) = \#_G(S)$ if S is regular and $S^{\mathcal{J}}$ is combinatorial and for other classes of semigroups. The details will be presented elsewhere. Tilson's Thesis [10] contains further results and techniques for showing $\#_I(S) = \#_G(S)$.

Added in proof: For a continuation of this paper see the forthcoming paper "Improved Lower Bounds for the Complexity of Finite Semigroups" by the authors, to appear in Advances in Mathematics, in which an example S is given with $\#_I(S) \lneqq \#_G(S)$. However, then stronger lower bounds for complexity will be derived in this future paper.

References

[1] K.Krohn, J.Rhodes and B.Tilson, Lectures on finite semigroups, chapters 1, 5–9 (chapter 6 with M.A.Arbib), in: The algebraic theory of machines, languages and semigroups, ed. M.A.Arbib (Academic Press, 1968).

[2] A.H.Clifford and G.B.Preston, Algebraic theory of semigroups, vol. 1, Math. Surveys no.7 (Am. Math. Soc., Providence, R.I., 1962).

[3] K.Krohn, R.Mateosian and J.Rhodes, Methods of the algebraic theory of machines, I. J.Computer and System Sciences 1 (1967) 55–85.

[4] P.H.Zieger, Cascade synthesis of finite-state machines, Information andControl 10 (1967) 419–433.

[5] J.Rhodes, A homomorphism theorem for finite semigroups, Math. System Theory 1 (1967) 289–304.

[6] K.Krohn and J.Rhodes, Complexity of finite semigroups, Ann. Math. 88 (1968) 128–160.

[7] K.Krohn, R.Mateosian and J.Rhodes, Complexity of ideals in finite semigroups and finite state machines, Math. System Theory 1 (1967) 59–66, and erratum 1 (1967) 373.

[8] J.Rhodes, Some results on finite semigroups, J. Algebra 4 (1966) 472–504.

[9] D.Allen, Relations between the local and global structure of finite semigroups, Ph.D. thesis, University of California, Berkeley (1968).

[10] B.Tilson, Ph.D. thesis, University of California, Berkeley, Advan. Math., to be published.

[11] Y.Zalcstein, Ph.D. thesis, University of California, Berkeley, Advan. Math., to be published.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 1 (1971) pp. 97–140.

COHERENCE IN CLOSED CATEGORIES

G.M. KELLY
The University of New South Wales, Australia

and

S. MACLANE
The University of Chicago, Ill., USA

Received 20 February 1970

§ 1. Introduction

For the purposes of this paper we understand by a *closed category* the following collection of data:

(i) a category $\underline{V}$;

(ii) functors $\otimes$: $\underline{V} \times \underline{V} \to \underline{V}$ and $[\ ,\ ]$: $\underline{V}^{\mathrm{op}} \times \underline{V} \to \underline{V}$;

(iii) an object I of $\underline{V}$;

(iv) natural isomorphisms

$$a = a_{ABC}\colon (A \otimes B) \otimes C \to A \otimes (B \otimes C)\,,$$

$$b = b_A : A \otimes I \to A\ ,$$

$$c = c_{AB}\colon A \otimes B \to B \otimes A\ ;$$

(v) natural transformations (in the generalized sense of [1])

$$d = d_{AB}\colon A \to [B, A \otimes B]\ ,$$

$$e = e_{AB}\colon [A, B] \otimes A \to B\ .$$

The axioms to be satisfied by these data are that, for all $A, B, C, D \in \underline{V}$, the following diagrams should commute:

C1

$$\begin{tikzcd}
((A \otimes B) \otimes C) \otimes D \arrow[r, "a"] \arrow[d, "a \otimes 1"'] & (A \otimes B) \otimes (C \otimes D) \arrow[r, "a"] & A \otimes (B \otimes (C \otimes D)) \\
(A \otimes (B \otimes C)) \otimes D \arrow[rr, "a"'] & & A \otimes ((B \otimes C) \otimes D) \arrow[u, "1 \otimes a"']
\end{tikzcd}$$

C2

$$\begin{tikzcd}
(A \otimes B) \otimes I \arrow[rr, "a"] \arrow[dr, "b"'] & & A \otimes (B \otimes I) \arrow[dl, "1 \otimes b"] \\
& A \otimes B &
\end{tikzcd}$$

C3

$$\begin{tikzcd}
A \otimes B \arrow[r, "c"] \arrow[dr, "1"'] & B \otimes A \arrow[d, "c"] \\
& A \otimes B
\end{tikzcd}$$

C4

$$\begin{tikzcd}
(A \otimes B) \otimes C \arrow[r, "a"] \arrow[d, "c \otimes 1"'] & A \otimes (B \otimes C) \arrow[r, "c"] & (B \otimes C) \otimes A \arrow[d, "a"] \\
(B \otimes A) \otimes C \arrow[r, "a"'] & B \otimes (A \otimes C) \arrow[r, "1 \otimes c"'] & B \otimes (C \otimes A)
\end{tikzcd}$$

C5

$$\begin{tikzcd}
{[B, A]} \arrow[r, "d"] \arrow[dr, "1"'] & {[B, [B, A] \otimes B]} \arrow[d, "{[1, e]}"] \\
& {[B, A]}
\end{tikzcd}$$

C6

$$\begin{tikzcd}
A \otimes B \arrow[r, "d \otimes 1"] \arrow[dr, "1"'] & {[B, A \otimes B] \otimes B} \arrow[d, "e"] \\
& A \otimes B
\end{tikzcd}$$

Such a closed category, which we denote by the single letter $\underline{V}$, is not essentially different from what was called in [2] a "symmetric monoidal closed category". In particular we have a natural isomorphism

$$\pi : \underline{V}(A \otimes B, C) \to \underline{V}(A, [B, C])$$

where $\pi(f)$ is the composite

$$(1.1) \qquad A \xrightarrow[d]{} [B, A \otimes B] \xrightarrow[[1,f]]{} [B, C]$$

and $\pi^{-1}(g)$ is the composite

$$(1.2) \qquad A \otimes B \xrightarrow[g \otimes 1]{} [B, C] \otimes B \xrightarrow[e]{} C;$$

indeed the commutativity of C5 and C6 is exactly the condition that the natural transformations π and π^{-1} defined by (1.1) and (1.2) should be mutually inverse.

If we omit [,], d, and e from the data and C5 and C6 from the axioms, we obtain the description of what we shall call a *monoidal category*. (This was called a "symmetric monoidal category" in [2], but we shall consider no other kind.) The axioms C1–C4 are exactly (see [9] and [5]) what is needed to ensure that the natural isomorphisms a, b, c are *coherent* in the sense of [9]. Roughly speaking, this means that any diagram will commute if (as in the diagrams C1–C4) each arrow is a natural isomorphism manufactured from $1, a, b, c, a^{-1}, b^{-1}, c^{-1}$ by taking repeated $\otimes$-products. Another example of such a diagram would be

$$\begin{tikzcd}
(I \otimes A) \otimes B \arrow[rr, "a"] \arrow[d, "c \otimes 1"'] & & I \otimes (A \otimes B) \arrow[d, "c"] \\
(A \otimes I) \otimes B \arrow[r, "b \otimes 1"'] & A \otimes B \arrow[r, "b^{-1}"'] & (A \otimes B) \otimes I
\end{tikzcd}$$

Note that coherence asserts equality of *natural transformations*, and not of morphisms in $\underline{V}$ except insofar as these are components of natural transformations; thus it does not assert that $c: A \otimes A \to A \otimes A$ and $1: A \otimes A \to A \otimes A$ coincide, these being components of quite different natural transformations $c: A \otimes B \to B \otimes A$ and $1: A \otimes B \to A \otimes B$.

The question naturally arises whether the analogous coherence result holds for a *closed* category: does a diagram commute if each arrow is a natural transformation manufactured from $1, a, b, c, a^{-1}, b^{-1}, c^{-1}, d, e$ by the use of $\otimes$ and [,]? Evidence that *something* of this kind is true was provided by the partial results in this direction due to Epstein [3] (cf. also MacDonald [8]), and by the mass of diagrams proved

to be commutative in [2]. Nevertheless the answer to the question as asked is negative. Write $k_A : A \to [[A, I], I]$ for the natural transformation given by the composite

$$A \xrightarrow[d]{} [[A, I], A \otimes [A, I]] \xrightarrow[[1,c]]{} [[A, I], [A, I] \otimes A] \xrightarrow[[1,e]]{} [[A, I], I] \; ;$$

then it is easy to see that the diagram

$$\begin{tikzcd}[A, I] \arrow[r, "k_{[A, I]}"] \arrow[dr, "1"'] & [[[A, I], I], I] \arrow[d, "{[k_A, 1]}"] \\ & [A, I]\end{tikzcd} \tag{1.3}$$

commutes; however the diagram

$$\begin{tikzcd}[[[A, I], I], I] \arrow[r, "{[k_A, 1]}"] \arrow[dr, "1"'] & [A, I] \arrow[d, "k_{[A, I]}"] \\ & [[[A, I], I], I]\end{tikzcd} \tag{1.4}$$

does not commute in general. For if (1.4) commuted as well as (1.3), $k_{[A, I]}$ would be an isomorphism; but this is not so when $\underline{V}$ is the category of real vector spaces with the usual $\otimes$ and $[\ ,\ ]$, in which case k_A is the usual embedding of a vector space into its double dual.

It is the chief purpose of the present paper to show that we do get a coherence result of the desired kind provided that we impose a restriction on the functors which form the "vertices" of the diagram: in the formation of these functors we must never write $[T, S]$ where S (like I in the example above) is a *constant* functor, unless T too is a constant functor. Both of the diagrams (1.3) and (1.4), then, escape this modified coherence result, as they stand; but (1.3) can equally well be written with a variable B in place of I, and then the result applies. Diagram (1.4) ceases to make sense, *as a diagram of natural transformations*, if we replace I by a variable B, and in fact as we have seen does not commute in general; but we could replace A in (1.4) by the constant I, and then our result applies and in this special case (1.4) commutes. The full statement of our results is given in §2 below.

The method we have used is inspired by the work of Lambek [6, 7], who deals

with a similar problem for a structure closely related to a closed category but differing from it in certain essential ways. (In particular, Lambek's structures lack the "symmetry isomorphism" c, and this does seem to make an essential difference.) From his work we have learnt the possibility of replacing *composition* of morphisms in a closed category by other processes of combination more adapted to proofs by induction. By his own account, Lambek himself came to recognize this possibility by generalizing the work of Gentzen, whose scheme for eliminating the "cut" in certain logical systems (see [4]) is essentially a special case of the above elimination of composition. An essential step in our §6 below, the proof that what we there call "constructible morphisms" are closed under composition, does not yield to a direct inductive argument – one must go round about and prove instead our Proposition 6.4; and this trick too we learnt from Lambek's work. These essential insights leave us heavily in Lambek's debt. For the rest, however, our results differ considerably from those of Lambek, being expressed in the context of the generalized natural transformations introduced in [1], with which the reader is supposed to be familiar.

§ 2. Statement of results

For a particular closed category $\underline{V}$, the functors $T, S: \underline{V} \times \underline{V}^{op} \times \underline{V} \to \underline{V}$, given by $T(A, B, C) = A \otimes [B, C]$ and $S(A, B, C) = [[A, B], C]$, might fortuitously coincide; they do so, in fact, if $\underline{V}$ is the category with one object and one morphism (and also in less trival cases). Since for our inductive arguments it is essential that each such functor be assigned a *rank*, and since the above two functors are to have different ranks, it is clear that rank should be an attribute not of the functor as such but of its formal expression. We proceed to introduce these formal expressions under the name of *shapes*.

We define shapes, without reference to any particular closed category, by the following inductive rules:

S1 I is a shape.

S2 1 is a shape.

S3 If T and S are shapes there is a shape $T \otimes S$.

S4 If T and S are shapes there is a shape $[T, S]$.

Shapes, therefore, are formal expressions involving I, $1, \otimes$, and $[\ ,\ ]$, with parentheses where necessary; for instance $[1, I] \otimes [1, (1 \otimes I) \otimes 1]$ is a shape.

We define a *variable-set* to be a totally-ordered finite set X, provided with a function called *variance* from X to the two-element set {covariant, contravariant}. Define the *ordinal sum* $X \dotplus Y$ of two variable-sets X and Y to be the disjoint union $X + Y$ of X and Y, so ordered that X and Y retain their orders and that every $x \in X$ precedes every $y \in Y$, and with the variance of $t \in X \dotplus Y$ being its variance in X or in Y as the case may be. Define the *twisted sum* $X \tilde{\dotplus} Y$ to be the same totally-ordered set as $X \dotplus Y$, but with the variance of $t \in X \tilde{\dotplus} Y$ being its variance in Y when $t \in Y$ and the opposite of its variance in X when $t \in X$.

With each shape T is associated a variable-set $v(T)$ called the *set of variables* of T. This is defined inductively by the rules:

V1 $v(I)$ is the empty set.

V2 $v(1)$ is a chosen one-element set $\{*\}$, with $*$ covariant.

V3 $v(T \otimes S) = v(T) \dotplus v(S)$.

V4 $v([T, S]) = v(T) \tilde{\dotplus} v(S)$.

In many contexts it is convenient to suppose that $v(T)$, if it has n elements, is actually the set $\{1, 2, ..., n\}$. We can accomplish this under the above conventions if we take $\{*\}$ to be $\{1\}$, and if we agree that the disjoint union of $\{1, ..., n\}$ and $\{1, ..., m\}$ is $\{1, ..., n+m\}$ with the given sets embedded as the complementary sets $\{1, ..., n\}$ and $\{n+1, ..., n+m\}$. We also, however, want to speak of $v(T)$ and $v(S)$ as being disjoint complementary subsets of $v(T) + v(S)$; the reader will recognize that we are then speaking of the *images* of $v(T)$ and $v(S)$ in $v(T) + v(S)$.

If T and S are shapes we define a *graph* $\xi: T \to S$ to be a fixed-point-free involution on the disjoint union $v(T) + v(S)$, with the property that mates under ξ have opposite variances in the *twisted* sum $v(T) \tilde{\dotplus} v(S)$. Given graphs $\xi: T \to S$ and $\eta: S \to R$, we define a composite graph $\eta\xi: T \to R$ as follows: different elements

$x, y \in v(T) + v(R)$ are mates under $\eta\xi$ if and only if there is a sequence $x = t_0, t_1, \ldots, t_r = y$, with each $t_i \in v(T) + v(S) + v(R)$, such that, for each i, t_{i-1} and t_i are mates either under ξ or under η. Then $\eta\xi$ is indeed a graph, and this law of composition is associative. Moreover for each shape T there is an evident identity graph $1: T \to T$, so that shapes and graphs form a category $\underline{G}$.

Two graphs $\xi: T \to S$, $\eta: S \to R$ are said to be *compatible* if there is no sequence $t_1, t_2, \ldots, t_{2r}$ $(r \geq 1)$ of elements of $v(S)$ such that t_{2i-1} and t_{2i} are mates under ξ for $1 \leq i \leq r$, t_{2i} and t_{2i+1} are mates under η for $1 \leq i \leq r-1$, and t_{2r} and t_1 are mates under η.

The definitions of composition of graphs and of compatibility of graphs become more perspicuous if we consider a graph $\xi: T \to S$ to be a graph in the literal sense, with the disjoint union $v(T) + v(S)$ as its vertex-set, and with one edge (or *linkage*) joining each pair of mates under ξ. The linkages in the graph $\eta\xi$ are then what we get by following alternately the linkages of ξ and of η, ignoring any closed loops that may arise; and ξ and η are compatible when in fact no closed loops do arise. All this is treated in detail in [1].

If $\xi: T \to T'$ and $\eta: S \to S'$ are graphs, we can define a graph $\xi \otimes \eta: T \otimes S \to T' \otimes S'$ by taking the linkages in $v(T) + v(S) + v(T') + v(S')$ to be those of ξ together with those of η. Similarly we can define a graph $[\xi, \eta]: [T', S] \to [T, S']$. It is easy to verify that $\otimes$ and $[\ ,\ ]$ are thereby made into functors $\underline{G} \times \underline{G} \to \underline{G}$ and $\underline{G}^{\mathrm{op}} \times \underline{G} \to \underline{G}$ respectively.

For any shapes T, S, R there are evident graphs

$$\alpha: (T \otimes S) \otimes R \to T \otimes (S \otimes R),$$

$$\beta: T \otimes I \to T,$$

$$\gamma: T \otimes S \to S \otimes T,$$

$$\delta: T \to [S, T \otimes S],$$

$$\epsilon: [T, S] \otimes T \to S;$$

it is easy to verify that these are natural transformations (natural isomorphisms in the case of α, β, γ), and that $\underline{G}$ becomes a closed category if we take these as its a, b, c, d, e.

Now let $\underline{V}$ be any closed category. For each shape T, with $v(T) = \{i_1, \ldots, i_n\}$ say (as an *ordered* set), we define a functor $|T|: \underline{V}_{i_1} \times \underline{V}_{i_2} \times \ldots \times \underline{V}_{i_n} \to \underline{V}$, where $\underline{V}_{i_r}$ is $\underline{V}$ or $\underline{V}^{\mathrm{op}}$ according as i_r is covariant or contravariant in $v(T)$. (If $v(T)$ is empty then $n = 0$ and we understand $\underline{V}_{i_1} \times \ldots \times \underline{V}_{i_n}$ to mean the unit category $\underline{I}$ with one object and one morphism.) The inductive definition of $|T|$ is the following:

F1 $|I|$ is the constant functor $I: \underline{I} \to \underline{V}$.

F2 $|1|$ is the identity functor $1: \underline{V} \to \underline{V}$.

F3 $|T \otimes S|$ is the composite functor

$$\underline{V}_{i_1} \times \ldots \times \underline{V}_{i_n} \times \underline{V}_{j_1} \times \ldots \times \underline{V}_{j_m} \xrightarrow[|T| \times |S|]{} \underline{V} \times \underline{V} \xrightarrow[\otimes]{} \underline{V},$$

where $v(T) = \{i_1, \ldots, i_n\}$ and $v(S) = \{j_1, \ldots, j_m\}$.

F4 $|[T, S]|$ is the composite functor

$$\underline{V}^{\mathrm{op}}_{i_1} \times \ldots \times \underline{V}^{\mathrm{op}}_{i_n} \times \underline{V}_{j_1} \times \ldots \times \underline{V}_{j_m} \xrightarrow[|T|^{\mathrm{op}} \times |S|]{} \underline{V}^{\mathrm{op}} \times \underline{V} \xrightarrow[[\,,\,]]{} \underline{V},$$

where $v(T)$ and $v(S)$ are as in F3.

Let T and S be shapes, with $v(T) = \{i_1, \ldots, i_n\}$ and $v(S) = \{j_1, \ldots, j_m\}$. Then, as in [1], a *natural transformation* $f\colon |T| \to |S|$ consists of a graph $\xi\colon T \to S$, called the *graph* Γf of f, and morphisms

$$f(A_{i_1}, \ldots, A_{i_n}, A_{j_1}, \ldots, A_{j_m})\colon |T|(A_{i_1}, \ldots, A_{i_n}) \to |S|(A_{j_1}, \ldots, A_{j_m})$$

of $\underline{V}$, called the *components* of f; here $A_x = A_y$ whenever x and y are mates under ξ, and for each such pair of mates there is a *naturality condition* to be satisfied by these components. (In practice one suppresses, in writing the components of f, one of each pair of equal arguments in $f(A_{i_1}, \ldots, A_{i_n}, A_{j_1}, \ldots, A_{j_m})$, and one often writes the remaining arguments as subscripts. Thus one writes $e_{AB}\colon [A, B] \otimes A \to B$ or $e(A, B)$, and not $e(A, B, A, B)$.) If $g\colon |S| \to |R|$ is another natural transformation, of graph $\eta\colon S \to R$, and if η and ξ are *compatible*, we can define as in [1] a composite natural transformation $gf\colon |T| \to |R|$ of graph $\eta\xi\colon T \to R$; the component

$$(gf)(A_{i_1}, \ldots, A_{i_n}, A_{k_1}, \ldots, A_{k_l})\colon |T|(A_{i_1}, \ldots, A_{i_n}) \to |R|(A_{k_1}, \ldots, A_{k_l})$$

of gf is the composite of the components

$$f(A_{i_1}, \ldots, A_{i_n}, A_{j_1}, \ldots, A_{j_m})$$

and

$$g(A_{j_1}, \ldots, A_{j_m}, A_{k_1}, \ldots, A_{k_l}),$$

where $A_x = A_y$ if x and y are either mates under ξ or mates under η. In fact we can in the present circumstances define the composite gf, of graph $\eta\xi$, even when η and ξ are not compatible, for we have here a recourse not available in the more general situation of [1]: we define the components of gf just as above, setting $A_{j_r} = I$ for any $j_r \in v(S)$ which occurs in one of the closed loops. That the composite so formed is still natural is clear, as we have merely modified f and g by specializing some of the arguments before composing them as in [1]. This law of composition is associative, and there is an evident identity natural transformation $1\colon |T| \to |T|$ of graph $1\colon T \to T$.

We can define, therefore, a new category $\underline{N}(\underline{V})$ depending upon $\underline{V}$. The objects of $\underline{N}(\underline{V})$, like those of $\underline{G}$, are to be all the shapes; a morphism $f\colon T \to S$ in $\underline{N}(\underline{V})$ is to be a natural transformation $f\colon |T| \to |S|$, which we shall often call "a natural transformation $f\colon T \to S$"; and composition in $\underline{N}(\underline{V})$ is to be the above composition of natural transformations. We can call $\underline{N}(\underline{V})$ "the category of shapes and natural

transformations for $\underline{V}$"; and we shall often abbreviate $\underline{N}(\underline{V})$ to $\underline{N}$ when $\underline{V}$ is clear from the context. There is an evident functor $\Gamma: \underline{N} \to \underline{G}$ which is the identity on objects and which takes each natural transformation f to its graph Γf.

From natural transformations $f: T \to T'$ and $g: S \to S'$ of graphs ξ and η we get a natural transformation $f \otimes g: T \otimes S \to T' \otimes S'$ of graph $\xi \otimes \eta$ by taking the components of $f \otimes g$ to be the $\otimes$-products of the components of f and those of g. Similarly we get a natural transformation $[f, g]: [T', S] \to [T, S']$ of graph $[\xi, \eta]$. It is easy to verify that $\otimes$ and $[\ ,\ ]$ are thereby made into functors $\underline{N} \times \underline{N} \to \underline{N}$ and $\underline{N}^{\mathrm{op}} \times \underline{N} \to \underline{N}$; clearly $\Gamma: \underline{N} \to \underline{G}$ commutes with $\otimes$ and $[\ ,\ ]$.

For any shapes T, S, R we get a natural transformation $a_{TSR}: (T \otimes S) \otimes R \to T \otimes (S \otimes R)$ of graph $\alpha_{TSR}: (T \otimes S) \otimes R \to T \otimes (S \otimes R)$ by taking the component

$$a_{TSR}(A_{i_1}, \ldots, A_{i_n}, A_{j_1}, \ldots, A_{j_m}, A_{k_1}, \ldots, A_{k_l})$$

of a_{TSR} to be the component

$$a(|T|(A_{i_1}, \ldots, A_{i_n}), |S|(A_{j_1}, \ldots, A_{j_m}), |R|(A_{k_1}, \ldots, A_{k_l}))$$

of a. Then it follows easily that the morphism a_{TSR} of $\underline{N}$ is the (T, S, R)-component of a natural isomorphism between the functors $(-\otimes-)\otimes-$ and $-\otimes(-\otimes-)$ of $\underline{N} \times \underline{N} \times \underline{N}$ into $\underline{N}$. This natural isomorphism we again call a, and we often write a. $(T \otimes S) \otimes R \to T \otimes (S \otimes R)$, abbreviating as usual a_{TSR} to a. In the same way we define natural isomorphisms $b: T \otimes I \to T$, $c: T \otimes S \to S \otimes T$ and natural transformations $d: T \to [S, T \otimes S]$, $e: [T, S] \otimes T \to S$, of respective graphs $\beta, \gamma, \delta, \epsilon$; and we verify that a, b, c, d, e give to $\underline{N}$ the structure of a closed category.

We now have closed categories $\underline{N} = \underline{N}(\underline{V})$ and $\underline{G}$, and a functor $\Gamma: \underline{N} \to \underline{G}$ which is the identity on objects, which commutes with $\otimes$ and $[\ ,\ ]$, and which sends a, b, c, d, e to $\alpha, \beta, \gamma, \delta, \epsilon$. In order to make statements that will embrace at once the closed categories $\underline{N}$ and $\underline{G}$, we shall suppose, throughout this paper, that $\underline{H}$ is some closed category with the same objects as $\underline{G}$, and that $\Gamma: \underline{H} \to \underline{G}$ is a functor which is the identity on objects, which commutes with $\otimes$ and $[\ ,\ ]$, and which sends a, b, c, d, e to $\alpha, \beta, \gamma, \delta, \epsilon$. The cases of interest are that where $\underline{H} = \underline{N}(\underline{V})$ and Γ is as above, and that where $\underline{H} = \underline{G}$ and $\Gamma = 1$.

Given any such $\underline{H}$, we define a subcategory of $\underline{H}$, whose objects are all shapes, and whose morphisms shall be called the *allowable* morphisms of $\underline{H}$. These are to be the smallest class of morphisms of $\underline{H}$ satisfying the following five conditions (in which $T, S, R, \ldots$ denote arbitrary shapes):

AM1. For any T, S, R each of the following morphisms is in the class:

$$\begin{aligned} 1 &: T \to T \\ a &: (T \otimes S) \otimes R \to T \otimes (S \otimes R), \\ a^{-1} &: T \otimes (S \otimes R) \to (T \otimes S) \otimes R, \\ b &: T \otimes I \to T, \\ b^{-1} &: T \to T \otimes I, \\ c &: T \otimes S \to S \otimes T. \end{aligned}$$

AM2. For any T, S each of the following morphisms is in the class:

$$d\colon\ T \to [S, T \otimes S]\ ,$$
$$e\colon\ [T, S] \otimes T \to S.$$

AM3. If $f\colon T \to T'$ and $g\colon S \to S'$ are in the class so is $f \otimes g\colon T \otimes S \to T' \otimes S'$.

AM4. If $f\colon T \to T'$ and $g\colon S \to S'$ are in the class so is $[f, g]\colon [T', S] \to [T, S']$.

AM5. If $f\colon T \to S$ and $g\colon S \to R$ are in the class so is $gf\colon T \to R$.

The allowable morphisms of $\underline{G}$ are called the *allowable graphs*, and those of $\underline{N}(\underline{V})$ are called the *allowable natural transformations* (for $\underline{V}$). It is evident that the functor $\Gamma\colon \underline{N} \to \underline{G}$ takes allowable natural transformations to allowable graphs, since those natural transformations $f \in \underline{N}$ for which Γf is allowable clearly satisfy AM1–AM5.

The first two of our principal results deal with the case $\underline{H} = \underline{G}$, and are:

Theorem 2.1. *There is an algorithm for deciding whether a graph $\xi\colon T \to S$ is allowable.*

Theorem 2.2. *If the graphs $\xi\colon T \to S$, $\eta\colon S \to R$ are allowable, they are compatible.*

The proofs will be given in §7 and in §6 respectively.

Since we shall be interested only in *allowable* natural transformations, we see from Theorem 2.2 that there was no real need to introduce the composition of incompatible ones; it was merely a convenience so that $\underline{N}$ could be described as a category. Our third principal result is:

Theorem 2.3. *Let $\underline{V}$ be any closed category. If $\xi\colon T \to S$ is an allowable graph, there is in $\underline{N}(\underline{V})$ at least one allowable natural transformation $f\colon T \to S$ of graph ξ.*

Proof. Those allowable graphs ξ which are images under Γ of allowable natural transformations satisfy AM1–AM5, and therefore constitute the totality of allowable graphs.

For our final main result we pick out a subset of the shapes called the *proper* shapes. Call a shape T *constant* if its set of variables $v(T)$ is empty. Then the proper shapes are defined inductively by:

PS1 $\quad I$ is a proper shape.

PS2 $\quad 1$ is a proper shape.

PS3 $\quad$ If T and S are proper shapes so is $T \otimes S$.

PS4 $\quad$ If T and S are proper shapes so is $[T, S]$, *unless* S is constant and T is not constant.

Our final principal result then is:

Theorem 2.4. *Let $\underline{V}$ be any closed category, and let $f, f': T \to S$ be two allowable natural transformations in $\underline{N}(\underline{V})$ with the same graph $\xi = \Gamma f = \Gamma f'$. Then, provided that the shapes T and S are proper, we have $f = f'$.*

The proof will be given in §7.

§3. The monoidal case

We are going to build on the known coherence theorem for the monoidal case, proved in [9] and simplified a little in [5]. The purpose of this section is to re-state this result in terms entirely analogous to those used in §2 above, so that it is easily available for our use.

We have seen that we get the description of a monoidal category from that of a closed category by omitting the data $[\ ,\ ]$, d, e and the axioms C5 and C6. All the concepts introduced in §2 have analogues in the monoidal case, as follows.

The shapes we need here are those defined by the inductive rules S1, S2, S3 of §2, omitting S4; we call these the *integral shapes* (for it is reasonable to think of $\otimes$ as a kind of multiplication, and of $[\ ,\ ]$ as a kind of division). For integral T the rules V1, V2, V3 suffice to describe the set of variables $v(T)$; clearly each element of $v(T)$ is covariant. Because of this, a pair of mates under a graph $\xi\colon T \to S$, where T and S are integral, consists of an element of $v(T)$ and an element of $v(S)$; thus we may identify the graph ξ with the corresponding bijection of $v(T)$ onto $v(S)$. It is especially for integral T (where there are no complications of variance) that it is convenient to identify $v(T)$, when it has n elements, with the ordered set $\{1, 2, ..., n\}$; and we shall do so freely. The integral shapes and the graphs connecting them form a full subcategory $\underline{G}_0$ of $\underline{G}$; we can look upon $\otimes$ as a functor $\underline{G}_0 \otimes \underline{G}_0 \to \underline{G}_0$, and the graphs $\alpha\colon (T \otimes S) \otimes R \to T \otimes (S \otimes R)$, $\beta\colon T \otimes I \to T$, $\gamma\colon T \otimes S \to S \otimes T$ turn $\underline{G}_0$ into a monoidal category.

If $\underline{V}$ is any monoidal category, each integral shape T determines a functor $|T|\colon \underline{V} \times ... \times \underline{V} \to \underline{V}$ by the rules F1, F2, F3 of §2. Since we can again speak of a natural transformation $f\colon |T| \to |S|$ of graph $\xi\colon T \to S$, we have a category $\underline{N}_0(\underline{V})$, whose objects are the integral shapes and whose morphisms $f\colon T \to S$ are the natural transformations $f\colon |T| \to |S|$ of arbitrary graph. Like the category $\underline{N}(\underline{V})$ of §2, $\underline{N}_0(\underline{V})$ becomes a monoidal category with the obvious definitions of $f \otimes g$ and of a, b, c; and there is a functor $\Gamma\colon \underline{N}_0(\underline{V}) \to \underline{G}_0$ which is the identity on objects and which sends each natural transformation to its graph. The functor Γ commutes with $\otimes$ and sends a, b, c, to α, β, γ.

In this monoidal case we shall need to compare the $\underline{N}_0(\underline{V})$'s for different monoidal categories $\underline{V}$. If $\underline{V}$ and $\underline{V}'$ are monoidal categories, a *strict monoidal functor* $\Delta\colon \underline{V} \to \underline{V}'$ shall mean a functor that commutes with $\otimes$ and for which $\Delta a = a'$, $\Delta b = b'$, and $\Delta c = c'$ (where, for example, this last assertion means that $\Delta c_{A,B} = c'_{\Delta A, \Delta B}$). In particular, $\Gamma\colon \underline{N}_0(\underline{V}) \to \underline{G}_0$ is a strict monoidal functor. It is easily seen that a strict monoidal functor $\Delta\colon \underline{V} \to \underline{V}'$ induces a strict monoidal functor $\underline{N}_0(\Delta)\colon \underline{N}_0(\underline{V}) \to \underline{N}_0(\underline{V}')$, which is the identity on objects and which sends the natural transformation $f\colon T \to S$ to the natural transformation whose components are the images under Δ of those of f. It is further clear that the composite of $\Gamma'\colon \underline{N}_0(\underline{V}') \to \underline{G}_0$ with $\underline{N}_0(\Delta)$ is $\Gamma\colon \underline{N}_0(\underline{V}) \to \underline{G}_0$.

For any monoidal category $\underline{V}$ we define the *central* morphisms of $\underline{V}$ to be the smallest class of morphisms of $\underline{V}$ satisfying the conditions AM1, AM3, and AM5 of

§2, where $T, S, R, \ldots$ now denote arbitrary objects of $\underline{V}$; since the isomorphisms of $\underline{V}$ satisfy AM1, AM3 and AM5, every central morphism is an isomorphism. These central morphisms constitute a subcategory Cent $\underline{V}$ of $\underline{V}$ with the same objects as $\underline{V}$; clearly Cent $\underline{V}$ is itself a monoidal category, and the inclusion Cent $\underline{V} \to \underline{V}$ is a strict monoidal functor. It is clear that any strict monoidal functor $\Delta : \underline{V} \to \underline{V}'$ carries central morphisms of $\underline{V}$ into central morphisms of $\underline{V}'$.

The analogue of Theorem 2.2 for the monoidal case is trivially true, for any graphs $\xi: T \to S$ and $\eta: S \to R$ are clearly compatible when T, S and R are integral. The analogues of Theorems 2.1, 2.3 and 2.4 are contained in the following result, which expresses essentially what was proved in [9]:

Theorem 3.1. *Let $\underline{V}$ be any monoidal category. If T and S are integral shapes, then any graph $\xi: T \to S$ is central in $\underline{G}_0$, and there is in* Cent $\underline{N}_0(\underline{V})$ *one and only one natural transformation $f: T \to S$ of graph ξ.*

In other words we have Cent $\underline{N}_0(\underline{V}) \cong$ Cent $\underline{G}_0 = \underline{G}_0$. We shall write $|\xi|_{\underline{V}}: T \to S$ for the unique morphism of Cent $\underline{N}_0(\underline{V})$ with $\Gamma|\xi|_{\underline{V}} = \xi$. It is immediate that $|\eta\xi|_{\underline{V}} = |\eta|_{\underline{V}}|\xi|_{\underline{V}}$, and that $|\xi \otimes \eta|_{\underline{V}} = |\xi|_{\underline{V}} \otimes |\eta|_{\underline{V}}$; and further that $|\alpha|_{\underline{V}} = a$, $|\beta|_{\underline{V}} = b$, $|\gamma|_{\underline{V}} = c$. Moreover if $\Delta: \underline{V} \to \underline{V}'$ is a strict monoidal functor, it is clear that $\underline{N}_0(\Delta)$ carries $|\xi|_{\underline{V}}$ to $|\xi|_{\underline{V}'}$.

§4. Central morphisms in $\underline{N}(\underline{V})$ and in $\underline{G}$

This section will use Theorem 3.1 to handle, for a closed category $\underline{V}$, that part of the coherence problem involving only a, b and c. In other words, we shall deal here with the *central* morphisms of $\underline{N}(\underline{V})$. These, like the composite

$$(T\otimes [S,R])\otimes I \xrightarrow[a]{} T\otimes ([S,R]\otimes I) \xrightarrow[1\otimes b]{} T\otimes [S,R]\ ,$$

involve in general non-integral shapes. We bring them within the ambit of Theorem 3.1 by showing that the central morphisms of $\underline{N}(\underline{V})$ and of $\underline{G}$ admit an alternative description: they arise from the morphisms of $\underline{G}_0$ by the substitution of "$\otimes$-irreducible" or "prime" shapes for the variables.

We suppose then that $\underline{V}$ is a closed category, and as in §2 we use $\underline{H}$ to denote either $\underline{N}(\underline{V})$ or $\underline{G}$, with $\Gamma: \underline{H} \to \underline{G}$ sending f to its graph in the first case and being the identity in the second case. Since $\underline{H}$, being a closed category, is a monoidal category, we can speak as in §3 of the central morphisms of $\underline{H}$; it is immediate from the definition of these that they are a subset of the allowable morphisms of $\underline{H}$. Since $\Gamma: \underline{N}(\underline{V}) \to \underline{G}$ is a strict monoidal functor, it takes a central morphism of $\underline{N}(\underline{V})$ (which we shall call a *central natural transformation*) to a central morphism of $\underline{G}$ (which we shall call a *central graph*). As $\underline{V}$ will be fixed, we shall abbreviate $\underline{N}(\underline{V})$ to $\underline{N}$.

If P is any integral shape we have as in §3, since $\underline{H}$ is a monoidal category, a functor $|P|: \underline{H}\times \dots \times \underline{H} \to \underline{H}$. Thus for arbitrary shapes $X_1, \dots, X_n$ (where n is the number of elements of $v(P)$) we get a shape $|P|(X_1, \dots, X_n)$, and for arbitrary morphisms $f_i: X_i \to X_i'$ in $\underline{H}$ we get a morphism $|P|(f_1, \dots, f_n): |P|(X_1, \dots, X_n) \to |P|(X_1', \dots, X_n')$ in $\underline{H}$. It is evident that $|P|(X_1, \dots, X_n)$ is the same shape whether we take $\underline{H}$ to be $\underline{N}$ or $\underline{G}$, and that $\Gamma(|P|(f_1, \dots, f_n)) = |P|(\Gamma f_1, \dots, \Gamma f_n)$.

Now let P, Q be integral shapes. A graph $\xi: P \to Q$ may be identified with a bijection of $v(P)$ onto $v(Q)$ and hence, if $v(P)$ and $v(Q)$ have n elements, with a permutation ξ of $\{1, \dots, n\}$. As in §3 we have a unique $|\xi|_{\underline{H}}: P \to Q$ in Cent $\underline{N}_0(\underline{H})$ of graph ξ. We can write its typical component as

$$(4.1)\qquad |\xi|_{\underline{H}}(X_1, \dots, X_n): |P|(X_{\xi 1}, \dots, X_{\xi n}) \to |Q|(X_1, \dots, X_n)\ ;$$

it is a morphism of $\underline{H}$.

Proposition 4.1. *For any graph $\xi: P \to Q$ between integral shapes P, Q and for any shapes $X_1, \dots, X_n$ the morphism* (4.1) *of $\underline{H}$ is central.*

Proof. Consider the family of all those graphs ξ in $\underline{G}_0$ for which (4.1) is indeed central in $\underline{H}$ for all $X_1, \dots, X_n$; it suffices to show that this family satisfies AM1, AM3 and AM5, for then it contains Cent $\underline{G}_0$ which, by Theorem 3.1, is all of $\underline{G}_0$. Now this family satisfies AM1 because $|\alpha|_{\underline{H}} = a$, etc.; it satisfies AM3 because

the components of $|\xi\otimes\eta|_{\underline{H}} = |\xi|_{\underline{H}}\otimes|\eta|_{\underline{H}}$ are the tensor products of the components of $|\xi|_{\underline{H}}$ and those of $|\eta|_{\underline{H}}$; and it satisfies AM5 because the components of $|\eta\xi|_{\underline{H}} = |\eta|_{\underline{H}}|\xi|_{\underline{H}}$ are the composites of certain components of $|\eta|_{\underline{H}}$ and of $|\xi|_{\underline{H}}$.

Since, as we saw in §3, $\underline{N}_0(\Gamma)$ takes $|\xi|_{\underline{N}}$ to $|\xi|_{\underline{G}}$, it follows from the definition in §3 of $\underline{N}_0(\Gamma)$ that

$$\Gamma(|\xi|_{\underline{N}}(X_1, ..., X_n)) = |\xi|_{\underline{G}}(X_1, ..., X_n)\,. \tag{4.2}$$

It is easy to calculate $|\xi|_{\underline{G}}(X_1, ..., X_n)$. First, it is clear by induction that the variable-set $v(|P|(X_1, ..., X_n))$ is $v(X_1) \dotplus ... \dotplus v(X_n)$.

Proposition 4.2. *The graph*

$$|\xi|_{\underline{G}}(X_1, ..., X_n)\colon |P|(X_{\xi 1}, ..., X_{\xi n}) \to |Q|(X_1, ..., X_n)$$

is the involution on the set

$$v(X_{\xi 1}) + ... + v(X_{\xi n}) + v(X_1) + ... + v(X_n)$$

corresponding to the evident bijection induced by ξ *of*

$$v(X_{\xi 1}) + ... + v(X_{\xi n}) \quad \textit{with} \quad v(X_1) + ... + v(X_n)\,.$$

Proof. Again it suffices to show that the family of those ξ in $\underline{G}_0$ for which this is true satisfies AM1, AM3, and AM5; the verifications are immediate.

We now proceed to show that all the central morphisms of $\underline{H}$ are obtainable in the form (4.1). Define the *prime* shapes to be the shape 1 and all shapes of the form $[T, S]$. It follows easily from the inductive definition of shapes that any shape T can be expressed *uniquely* in the form $T = |P|(X_1, ..., X_n)$ where P is an integral shape and $X_1, ..., X_n$ are prime shapes. We call this the *prime factorization* of T, and call $X_1, ..., X_n$ the list of *prime factors* of T. Note that n may be 0, so that this list may be empty; namely when T is a constant integral shape. In general, if T is an integral shape, its prime factorization is $T = |T|(1, 1, ..., 1)$. Observe that if the prime factorizations of T and S are $T = |P|(X_1, ..., X_n)$ and $S = |Q|(Y_1, ..., Y_m)$, then that of $T\otimes S$ is $|P\otimes Q|(X_1, ..., X_n, Y_1, ..., Y_m)$.

Proposition 4.3. *Let* $f\colon T\to S$ *be a central morphism of* $\underline{H}$, *and let the prime factorizations of* T *and of* S *be* $T = |P|(X_1, ..., X_n)$ *and* $S = |Q|(Y_1, ..., Y_m)$. *Then* $m = n$, *and there is a permutation* ξ *of* $\{1, ..., n\}$ *such that* $X_i = Y_{\xi i}$ *for each* i *and such that* $f = |\xi|_{\underline{H}}(Y_1, ..., Y_n)\colon |P|(Y_{\xi 1}, ..., Y_{\xi n}) \to |Q|(Y_1, ..., Y_n)$.

Proof. Consider the family of all those morphisms of $\underline{H}$ that *are* of the form

$$|\eta|_{\underline{H}}(Z_1, ..., Z_r)\colon |J|(Z_{\eta 1}, ..., Z_{\eta r}) \to |K|(Z_1, ..., Z_r)$$

for integral shapes J, K and prime shapes Z_i. It suffices to show that this family satisfies AM1, AM3 and AM5, and therefore contains all central morphisms of $\underline{H}$. When we advert to the relation between the prime factorization of $T \otimes S$ and those of T and of S, the verifications are immediate from the facts that $|\alpha|_{\underline{H}} = a$, etc., $|\eta \otimes \zeta|_{\underline{H}} = |\eta|_{\underline{H}} \otimes |\zeta|_{\underline{H}}$, and $|\zeta\eta|_{\underline{H}} = |\zeta|_{\underline{H}}\, |\eta|_{\underline{H}}$.

Since a shape T is integral exactly when its prime factors are all 1, and is constant exactly when its prime factors are all constant, we have

Corollary 4.4. *If $f: T \to S$ is a central morphism in $\underline{H}$ and if either one of T, S is integral (resp. constant), so is the other.*

In view of Proposition 4.2 we also have

Corollary 4.5. *If $\phi: T \to S$ is a central graph, each pair of mates under ϕ consists of an element of $v(T)$ and an element of $v(S)$.*

Returning to Proposition 4.3, we may observe that the permutation ξ therein is not in general uniquely determined by f. For instance if T and S are both $[I, I] \otimes [I, I]$, so that $P = Q = 1 \otimes 1$ and $X_1 = Y_1 = X_2 = Y_2 = [I, I]$, then it will follow from Proposition 4.8 below that $|\xi|_{\underline{H}}(Y_1, Y_2)$ is $1: T \to S$ for both permutations ξ of $\{1, 2\}$. However:

Proposition 4.6. *If in Proposition 4.3 permutations ξ and ξ' both satisfy the stated conditions, we have $\xi' = \lambda\xi$ where λ is a permutation of $\{1, ..., n\}$ for which $\lambda i \neq i$ implies that Y_i and $Y_{\lambda i}$ are equal constant shapes.*

Proof. Since $X_{\xi^{-1} i} = Y_i$ and also $X_{\xi^{-1} i} = Y_{\xi'\xi^{-1} i} = Y_{\lambda i}$ we have $Y_i = Y_{\lambda i}$. Since $|\xi|_{\underline{H}}(Y_1, ..., Y_n) = |\xi'|_{\underline{H}}(Y_1, ..., Y_n)$ we have by (4.2) that $|\xi|_{\underline{G}}(Y_1, ..., Y_n) = |\xi'|_{\underline{G}}(Y_1, ..., Y_n)$, and we conclude from Proposition 4.2 that $\xi' j = \xi j$ unless $v(Y_{\xi j})$ is empty; that is, $\lambda i = i$ unless Y_i is constant.

For the desired main result of this section, we need to show that the permutations λ of the type described in Proposition 4.6 are *exactly* those for which $|\lambda|_{\underline{H}}(Y_1, ..., Y_n) = 1$. First we prove:

Lemma 4.7. *If T is a constant shape there is an isomorphism $k_T: T \to I$ in $\underline{H}$ which, together with its inverse, is allowable.*

Proof. From the natural isomorphism

$$\underline{H}(A, I) \xrightarrow[\underline{H}(b,1)]{} \underline{H}(A \otimes I, I) \xrightarrow[\pi]{} \underline{H}(A, [I, I])$$

we deduce, by the Yoneda Lemma, the existence of an isomorphism $h: [I, I] \to I$.

Using (1.1) and (1.2) we find that h and h^{-1} are the respective composites

$$[I,I] \xrightarrow[b^{-1}]{} [I,I]\otimes I \xrightarrow[e]{} I, \qquad I \xrightarrow[d]{} [I,I\otimes I] \xrightarrow[[1,b]]{} [I,I],$$

so that both are allowable. We now define k_T inductively for constant shapes T by setting $k_I = 1$, by taking $k_{T\otimes S}$ to be the composite

$$T\otimes S \xrightarrow[k_T\otimes k_S]{} I\otimes I \xrightarrow[b]{} I,$$

and by taking $k_{[T,S]}$ to be the composite

$$[T,S] \xrightarrow[[k_T^{-1},k_S]]{} [I,I] \xrightarrow[h]{} I.$$

Proposition 4.8. *Let Q be an integral shape, Let $Y_1, \ldots, Y_n$ be prime shapes, and let λ be a permutation of $\{1, \ldots, n\}$ for which $\lambda i \neq i$ implies that Y_i and $Y_{\lambda i}$ are equal constant shapes. Then* $|\lambda|_{\underline{H}}(Y_1, \ldots, Y_n) = 1\colon |Q|(Y_{\lambda 1}, \ldots, Y_{\lambda n}) \to |Q|(Y_1, \ldots, Y_n)$.

Proof. We can express λ as a product of transpositions; since $|\mu\nu|_{\underline{H}} = |\mu|_{\underline{H}}|\nu|_{\underline{H}}$ we may suppose that λ is such a transposition. Replacing λ by a suitable conjugate $\mu\lambda\mu^{-1}$, we may suppose that λ is the transposition interchanging 1 and 2 and leaving fixed 3, ..., n, while $Y_1 = Y_2$ are equal constant shapes. In $\underline{G}_0$, Q is isomorphic to $(1\otimes 1)\otimes R$ for some integral R, and since $|\mu\otimes 1|_{\underline{H}} = |\mu|_{\underline{H}}\otimes 1$ we may suppose that Q is in fact the shape $1\otimes 1$. But then $|\lambda|_{\underline{H}} = c$, and it remains to prove that $c_{YY} = 1\colon Y\otimes Y \to Y\otimes Y$ if Y is a constant shape. Using the isomorphism k_Y of Lemma 4.7 we have by the naturality of c a commutative diagram

$$\begin{tikzcd}
Y\otimes Y \arrow[r, "c_{YY}"] \arrow[d, "k_Y\otimes k_Y"'] & Y\otimes Y \arrow[d, "k_Y\otimes k_Y"] \\
I\otimes I \arrow[r, "c_{II}"'] & I\otimes I
\end{tikzcd};$$

since $c_{II} = 1$ by Theorem 3.1, it follows that $c_{YY} = 1$.

Theorem 4.9. *Let $\underline{V}$ be a closed category. Then each central graph $\phi\colon T\to S$ in $\underline{G}$ is Γf for a unique central natural transformation $f\colon T\to S$ in $\underline{N}(\underline{V})$.*

Proof. Let $T = |P|(X_1, \ldots, X_n)$, $S = |Q|(Y_1, \ldots, Y_m)$ be the prime factorizations. Applying Proposition 4.3 with $\underline{H} = \underline{G}$, we conclude that $m = n$, and that for some permutation ξ we have $X_i = Y_{\xi i}$ and $\phi = |\xi|_{\underline{G}}(Y_1, \ldots, Y_n)$. Setting $f = |\xi|_{\underline{N}}(Y_1, \ldots, Y_n)$, which is central by Proposition 4.1, we see by (4.2) that $\Gamma f = \phi$, thus proving the existence of f.

To prove the uniqueness of f, let $f': T \to S$ be another central natural transformation with $\Gamma f' = \phi$. Applying Proposition 4.3 with $\underline{H} = \underline{N}$, we conclude that $f' = |\xi'|_{\underline{N}}(Y_1, ..., Y_n)$ for some permutation ξ' with $X_i = Y_{\xi' i}$. Now (4.2) gives $\phi = \Gamma f' = |\xi'|_{\underline{G}}(Y_1, ..., Y_n)$; and Proposition 4.6 with $\underline{H} = \underline{G}$ shows that $\lambda = \xi'\xi^{-1}$ has the properties described therein. We conclude from Proposition 4.8 with $\underline{H} = \underline{N}$ that $|\lambda|_{\underline{N}}(Y_1, ..., Y_n) = 1$. Thus, since $|\lambda\xi|_{\underline{N}} = |\lambda|_{\underline{N}} |\xi|_{\underline{N}}$, we have $|\xi'|_{\underline{N}}(Y_1, ..., Y_n) = |\xi|_{\underline{N}}(Y_1, ..., Y_n)$, or $f' = f$.

We conclude this section with two useful propositions that could in fact have been proved immediately after Proposition 4.3. In the situation of that proposition, we may call ξ the *association of the prime factors of T and of S*, and then call $Y_{\xi i}$ *the prime factor of S associated, via f, with the prime factor X_i of T*. This language is a little imprecise, because of the non-uniqueness of ξ; we sometimes have a *choice of association*. The statements of the results below allow for this choice.

Proposition 4.10. *Let $f: A \otimes B \to C \otimes D$ be a central morphism of $\underline{H}$. For some choice of association, let each prime factor of A, considered as a prime factor of $A \otimes B$, be associated via f with a prime factor of C. Then there are a shape E and central morphisms $g: A \otimes E \to C$ and $h: B \to E \otimes D$ such that f is the composite*

$$A \otimes B \xrightarrow[1 \otimes h]{} A \otimes (E \otimes D) \xrightarrow[a^{-1}]{} (A \otimes E) \xrightarrow[g \otimes 1]{} C \otimes D.$$

Proof. Let the prime factorizations be $A = |P|(X_1, ..., X_n)$, $B = |Q|(Y_1, ..., Y_m)$, $C = |R|(Z_1, ..., Z_l)$, $D = |S|(V_1, ..., V_k)$. We must have $n + m = l + k$, and the hypothesis of the proposition means that $f = |\xi|_{\underline{H}}(Z_1, ..., Z_l, V_1, ..., V_k)$ for some permutation ξ of $\{1, ..., n+m\}$ that maps the subset $\{1, ..., n\}$ into the subset $\{1, ..., l\}$. Let $j_1, ..., j_{l-n}$ be those elements of $\{1, ..., l\}$, in ascending order, that are not in the image under ξ of $\{1, ..., n\}$. Set $E = (Z_{j_1} \otimes Z_{j_2}) \otimes ... \otimes Z_{j_{l-n}}$; any way of inserting parentheses will do. Let ρ be the permutation of $\{1, ..., l\}$ given by $\rho i = \xi i$ for $i \leq n$, $\rho(n+i) = j_i$ for $i \leq l - n$. Let $\bar{\rho}$ be the permutation of $\{1, ..., n+m\}$ which is equal to ρ on $\{1, ..., l\}$ and which is the identity on $\{l+1, ..., n+m\}$. Let $\bar{\sigma}$ be the permutation $\bar{\rho}^{-1}\xi$ of $\{1, ..., n+m\}$; clearly $\bar{\sigma}$ is the identity on $\{1, ..., n\}$. Let σ be the permutation of $\{1, ..., m\}$ given by $\sigma i = \bar{\sigma}(n+i) - n$. Define g as $|\rho|_{\underline{H}}(Z_1, ..., Z_l)$ and h as $|\sigma|_{\underline{H}}(Z_{j_1}, ..., Z_{j_{l-n}}, V_1, ..., V_k)$. Then $1 \otimes h = |\bar{\sigma}|_{\underline{H}}(X_1, ..., X_n, Z_{j_1}, ..., Z_{j_{l-n}}, V_1, ..., V_k)$; $a^{-1} = |\alpha^{-1}|_{\underline{H}}(X_1, ..., X_n, Z_{j_1}, ..., Z_{j_{l-n}}, V_1, ..., V_k) = |1|_{\underline{H}}(X_1, ..., V_k)$; and $g \otimes 1 = |\bar{\rho}|_{\underline{H}}(Z_1, ..., Z_l, V_1, ..., V_k)$. It follows that $(g \otimes 1) a^{-1}(1 \otimes h) = |\bar{\rho}\bar{\sigma}|_{\underline{H}}(Z_1, ..., Z_l, V_1, ..., V_k) = |\xi|_{\underline{H}}(Z_1, ..., V_k) = f$, as required.

Proposition 4.11. *Let $f: [P, Q] \otimes B \to [R, S] \otimes D$ be a central morphism of $\underline{H}$, and for some choice of association let the prime factor $[P, Q]$ of $[P, Q] \otimes B$ be asso-*

ciated via f *with the prime factor* $[R, S]$ *of* $[R, S] \otimes D$. *Then* $P = R$, $Q = S$, *and there is a central morphism* $k\colon B \to D$ *such that* $f = 1 \otimes k : [P, Q] \otimes B \to [P, Q] \otimes D$.

Proof. Since associated prime factors must be equal, we have $P = R$ and $Q = S$. We apply Proposition 4.10 with $A = [P, Q]$ and $C = [R, S]$; in this case $E = I$, and g is clearly $b\colon [P, Q] \otimes I \to [P, Q]$. Writing k for the composite

$$B \xrightarrow[h]{} I \otimes D \xrightarrow[c]{} D \otimes I \xrightarrow[b]{} D,$$

it follows from Theorem 3.1 that $f = 1 \otimes k$.

§5. Processes of construction

In the next section we shall show that the allowable natural transformations and the allowable graphs can be classified by a numerical *rank*, and that those of higher rank can be built up from those of lower rank, modulo central ones, by the use of three simple processes now to be described.

We consider a closed category $\underline{H}$, which will in our applications be either $\underline{G}$ or $\underline{N}(\underline{V})$. The first process of construction is the formation of the tensor product $f \otimes g : A \otimes B \to C \otimes D$ of two given morphisms $f: A \to C$ and $g: B \to D$. Observe that

$$hf \otimes kg = (h \otimes k)(f \otimes g) \tag{5.1}$$

whenever hf and gk are defined. The second process of construction is the formation of the morphism $\pi(f): A \to [B, C]$ as in §1 from a given morphism $f: A \otimes B \to C$. Since π is natural we have commutativity in

$$\begin{tikzcd} A' \arrow[d, "g"'] \arrow[dr, "\pi(f(g \otimes 1))"] & \\ A \arrow[r, "\pi(f)"'] & {[B, C]} \end{tikzcd} \tag{5.2}$$

where $f(g \otimes 1)$ is the obvious composite $A' \otimes B \to A \otimes B \to C$. The third process of construction begins with morphisms $f: A \to B$ and $g: C \otimes D \to E$ and produces the composite

$$([B, C] \otimes A) \otimes D \xrightarrow[(1 \otimes f) \otimes 1]{} ([B, C] \otimes B) \otimes D \xrightarrow[e \otimes 1]{} C \otimes D \xrightarrow[g]{} E. \tag{5.3}$$

Rather than introduce a special symbol for the composite (5.3), we find it convenient to denote the composite

$$[B, C] \otimes A \xrightarrow[1 \otimes f]{} [B, C] \otimes B \xrightarrow[e]{} C \tag{5.4}$$

by $\langle f \rangle : [B, C] \otimes A \to C$, so that (5.3) may be written as $g(\langle f \rangle \otimes 1)$. The symbol $\langle f \rangle$ is of course ambiguous, inasmuch as the value of C must be understood from the context. It is clear that $\langle\ \rangle$ is natural, in the sense that, for

$$A' \xrightarrow[u]{} A \xrightarrow[f]{} B \xrightarrow[v]{} B', \quad C' \xrightarrow[w]{} C$$

we have commutativity in

$$(5.5)\qquad \begin{tikzcd} {[B,C]\otimes A} \arrow[r, "\langle f\rangle"] & C \\ {[B',C']\otimes A'} \arrow[u, "{[v,w]\otimes u}"] \arrow[r, "\langle vfu\rangle"'] & C' \arrow[u, "w"'] \end{tikzcd}$$

We need the following two lemmas giving connections between these processes.

Lemma 5.1. *Let $f: A\otimes B \to C$ and $g: D \to I$. Then the image under π of the composite*

$$(A\otimes D)\otimes B \xrightarrow[u]{} (A\otimes B)\otimes D \xrightarrow[f\otimes g]{} C\otimes I \xrightarrow[b]{} C ,$$

where u is the evident central morphism $a^{-1}(1\otimes c)a$, is the composite

$$A\otimes D \xrightarrow[\pi(f)\otimes g]{} [B,C]\otimes I \xrightarrow[b]{} [B,C] .$$

Proof. Set $\pi(f) = h$, so that $f = e(h\otimes 1)$ by the definition of π. Then

$$b(f\otimes g)u = b(e(h\otimes 1)\otimes g)u = b(e\otimes 1)((h\times 1)\otimes g)u ,$$

which is $ebu((h\otimes g)\otimes 1)$ by the naturality of b and of u. The bu in this last expression is, by Theorem 3.1, the unique central morphism $b\otimes 1: ([B,C]\otimes I)\otimes B \to [B,C]\otimes B$. Thus, by the definition of π again,

$$b(f\otimes g)u = e(b\otimes 1)((h\otimes g)\otimes 1) = \pi^{-1}(b(h\otimes g)) .$$

Lemma 5.2. *For $f: A\to B$ and $g: C\otimes B \to D$, we have*

$$g(1\otimes f) = \langle f\rangle(\pi(g)\otimes 1): C\otimes A \to D .$$

Proof. By (5.1) and the definition of π,

$$\langle f\rangle(\pi(g)\otimes 1) = e(1\otimes f)(\pi(g)\otimes 1)) = e(\pi(g)\otimes 1)(1\otimes f) = g(1\otimes f) .$$

The remainder of this section concerns compatibility of graphs, for the closed category $\underline{G}$. We mostly omit the proofs, which are entirely evident but tedious to put into words.

Lemma 5.3. *Let $\xi : Q \to R$, $\eta : R \to S$, $\zeta : S \to T$ be graphs in $\underline{G}$; then the following assertions are equivalent:*

(i) *ζ is compatible with η and $\zeta\eta$ is compatible with ξ;*

(ii) *η is compatible with ξ and ζ is compatible with $\eta\xi$.*

When the assertions of Lemma 5.3 are true, we say that the three graphs ξ, η, ζ are compatible. The concept clearly extends to any number of graphs $\xi_i : T_{i-1} \to T_i$.

Lemma 5.4. (a) *Graphs $\xi : R \to S$ and $\eta : S \to T$ are compatible if either is central.*

(b) *If two graphs $\xi : R \to S$ and $\eta : S \to T$ are compatible while the graphs $\rho : R' \to R$, $\sigma : S \to S'$, $\tau : T \to T'$ are central, then the graphs $\sigma\xi\rho : R' \to S'$ and $\tau\eta\sigma^{-1} : S' \to T'$ are compatible.*

Proof. Immediate from Corollary 4.5.

Lemma 5.5. *In the situation of (5.1) above, if $\underline{H} = \underline{G}$, $h \otimes k$ is compatible with $f \otimes g$ if and only if h is compatible with f and k compatible with g.*

Lemma 5.6. *In the situation of (5.2) above, if $\underline{H} = \underline{G}$, then $\pi(f)$ is compatible with g if and only if f is compatible with $g \otimes 1$.*

Lemma 5.7. *Let $f : A \to B$, $g : C \otimes D \to E$, $k : E \otimes F \to G$ in $\underline{G}$. If k is compatible with $g \otimes 1 : (C \otimes D) \otimes F \to E \otimes F$, it is compatible with $(g \otimes 1)((\langle f \rangle \otimes 1) \otimes 1)$: $(([B, C] \otimes A) \otimes D) \otimes F \to E \otimes F$.*

Lemma 5.8. *Let $f : A \to B$, $g : C \otimes D \to E$, $u : A' \to A$, $v : D' \to D$ in $\underline{G}$. If f is compatible with u and if g is compatible with $1 \otimes v$ then $g(\langle f \rangle \otimes 1) : ([B, C] \otimes A) \otimes D \to E$ is compatible with $(1 \otimes u) \otimes v : ([B, C] \otimes A') \otimes D' \to ([B, C] \otimes A) \otimes D$.*

Lemma 5.9. *Let $\underline{H} = \underline{G}$, let f and g be as in Lemma 5.2, and let $h : D \otimes E \to F$. Then if $h : D \otimes E \to F$. Then if $h : D \otimes E \to F$, $g \otimes 1 : (C \otimes B) \otimes E \to D \otimes E$, and $(1 \otimes f) \otimes 1 : (C \otimes A) \otimes E \to (C \otimes B) \otimes E$ are compatible, so are $h(\langle f \rangle \otimes 1) : ([B, D] \otimes A) \otimes E \to F$ and $(\pi(g) \otimes 1) \otimes 1 : (C \otimes A) \otimes E \to ([B, D] \otimes A) \otimes E$.*

§6. Constructibility of allowable morphisms

We place ourselves once again in the general situation $\Gamma: \underline{H} \to \underline{G}$ envisaged in §2 and §4; we recall that the cases of interest are $\underline{H} = \underline{N}(\underline{V})$ and $\underline{H} = \underline{G}$. The object of this section is to show that the allowable morphisms may be built up, modulo central morphisms, by the three processes described in §5. It is convenient to introduce the temporary name of *constructible* morphisms for those allowable morphisms that can be so built up; our aim is then to show that all allowable morphisms are constructible. We also give in this section the proof of Theorem 2.2.

We therefore define the *constructible* morphisms of $\underline{H}$ to be the smallest class of morphisms of $\underline{H}$ satisfying the following five conditions:

CM1 Every central morphism is in the class.

CM2 If $f: T \to S$ is in the class and if $u: T' \to T$ and $v: S \to S'$ are central then $vfu: T' \to S'$ is in the class.

CM3 If $f: A \to C$ and $g: B \to D$ are in the class so is $f \otimes g: A \otimes B \to C \otimes D$.

CM4 If $f: A \otimes B \to C$ is in the class so is $\pi(f): A \to [B, C]$.

CM5 If $f: A \to B$ and $g: C \otimes D \to E$ are in the class so is $g(\langle f \rangle \otimes 1)$: $([B, C] \otimes A) \otimes D \to E$.

In view of the definitions (1.1) of π and (5.4) of $\langle\ \rangle$, it is evident that the allowable morphisms satisfy CM1–CM5, so that the constructible morphisms are a subclass of the allowable ones.

We call an allowable morphism $f: T \to S$ in $\underline{H}$ *trivial* if both T and S are constant integral shapes.

Lemma 6.1. *A trivial constructible morphism in $\underline{H}$ is central.*

Proof. Consider the subclass of the constructible morphisms consisting of the following morphisms $f: T \to S$: if T and S are both constant integral shapes, f is to be central; otherwise, f is to be constructible. This subclass clearly satisfies CM1, CM4 and CM5. It satisfies CM2 because, by Corollary 4.4, if T' and S' are constant integral shapes so are T and S; and then vfu is central if f is. It satisfies CM3 because if $A \otimes B$ and $C \otimes D$ are constant integral shapes so are A, B, C and D; and then $f \otimes g$ is central if f and g are. Hence this subclass contains all constructible morphisms.

Proposition 6.2. *For each constructible $h: T \to S$ in $\underline{H}$, at least one of the following is true:*

(i) *h is central*

(ii) *h is of the form*

$$T \xrightarrow[x]{} A \otimes B \xrightarrow[f \otimes g]{} C \otimes D \xrightarrow[y]{} S$$

where f and g are constructible and non-trivial, and x and y are central.

(iii) *h is of the form*

$$T \xrightarrow[\pi(f)]{} [B,C] \xrightarrow[y]{} S$$

where f is constructible and y is central.

(iv) *h is of the form*

$$T \xrightarrow[x]{} ([B,C] \otimes A) \otimes D \xrightarrow[\langle f \rangle \otimes 1]{} C \otimes D \xrightarrow[g]{} S$$

where g and f are constructible and x is central.

Proof. Consider those constructible morphisms that *are* of one of the above forms (i)–(iv); we show that this class satisfies CM1–CM5 and therefore consists of all constructible morphisms. That it satisfies CM1, CM4 and CM5 is clear.

To see that CM2 is satisfied, let $u: T' \to T$ and $v: S \to S'$ be central. Then if h is central, so is vhu. If h is as in (ii) above, vhu is $(vy)(f \otimes g)(xu)$, which is of the same form. If h is as in (iii) above, vhu is $(vy)\,\pi(f(u \otimes 1))$, which is of the same form, $f(u \otimes 1)$ being constructible by CM2 since $u \otimes 1$ is central. If h is as in (iv) above, vhu is $(vg)(\langle f \rangle \otimes 1)(xu)$, which is of the same form, vg being constructible by CM2.

That CM3 is satisfied is clear unless f or g is trivial. If g is trivial it is central by Lemma 6.1. In this case B, I, D are constant integral shapes, and the empty graphs $B \to I$ and $I \to D$ give, by Proposition 4.1, central morphisms $u: B \to I$ and $v: I \to D$ in $\underline{H}$. Then by Theorem 4.9 we have $g = vu$. It follows at once from the naturality of b that $f \otimes g$ is then the composite

$$A \otimes B \xrightarrow[1 \otimes u]{} A \otimes I \xrightarrow[b]{} A \xrightarrow[f]{} C \xrightarrow[b^{-1}]{} C \otimes I \xrightarrow[1 \otimes v]{} C \otimes D;$$

since $(1 \otimes v)b^{-1}$ and $b(1 \otimes u)$ are central, and since CM2 is satisfied, this lies in the class because f does. Finally if f is trivial then, by the naturality of c, $f \otimes g$ is the composite

$$A \otimes B \xrightarrow[c]{} B \otimes A \xrightarrow[g \otimes f]{} D \otimes C \xrightarrow[c]{} C \otimes D,$$

which is in the class since CM2 is satisfied and since, by what we have just proved, $g \otimes f$ is in the class.

Remark. For brevity, morphisms h of the forms (ii), (iii), (iv) of Proposition 6.2 will be said to be respectively *of type* $\otimes$, *of type* π, and *of type* $\langle\ \rangle$.

For the purposes of our inductive proofs we introduce for each shape T a non-negative integer $r(T)$ called its *rank*, defined by the following inductive rules:

R1 $\qquad r(I) = 0$.

R2 $\quad r(1) = 1$.

R3 $\quad r(T \otimes S) = r(T) + r(S)$.

R4 $\quad r([T, S]) = r(T) + r(S) + 1$.

Note that $r(T) = 0$ if and only if T is a constant integral shape.

Lemma 6.3. *If $f\colon T \to S$ is central then $r(T) = r(S)$.*

Proof. Those central morphisms for which this is true clearly satisfy AM1, AM3 and AM5, and therefore constitute the totality of central morphisms.

The non-trivial step in the proof that all the allowable morphisms are constructible is the proof that the constructible morphisms are closed under composition. In fact, because of the exigencies of the inductive argument, we prove the variant of closure-under-composition given in Proposition 6.4 below. Moreover, because the same inductive argument applies, we prove at the same time the corresponding fact about compatibility, which will lead to a proof of Theorem 2.2.

Proposition 6.4. *If the morphisms $h\colon T \to S$ and $k\colon S \otimes U \to V$ of $\underline{H}$ are constructible, so is the composite morphism*

$$T \otimes U \xrightarrow[h \otimes 1]{} S \otimes U \xrightarrow[k]{} V.$$

Moreover, if $\underline{H} = \underline{G}$, the graphs k and $h \otimes 1$ are compatible.

Proof. The proof is by a double induction; we suppose the results to be true for all pairs of constructible morphisms $h'\colon T' \to S'$ and $k'\colon S' \otimes U' \to V'$ for which $r(T') + r(S') + r(U') + r(V') < r(T) + r(S) + r(U) + r(V)$; we also suppose them to be true for any pair h', k' for which $r(T') + r(S') + r(U') + r(V') = r(T) + r(S) + r(U) + r(V)$, provided that $r(T') + r(S') < r(T) + r(S)$.

By Proposition 6.2, each of h and k is central, or of type $\otimes$, or of type π, or of type $\langle\ \rangle$; we distinguish cases accordingly. We shall use Lemma 5.4, the Axiom CM2, and Lemma 6.3 freely without further explicit mention to "ignore" or to "absorb" central morphisms wherever convenient.

Case 1: either h or k is central. If h is central, so is $h \otimes 1$; the results follow from CM2 and from Lemma 5.4.

Case 2: h is of type $\langle\ \rangle$. Let h be $g(\langle f\rangle \otimes 1)x$ as in Proposition 6.2 (iv). Then the desired composite is

$$k(h \otimes 1) = k(g \otimes 1)((\langle f\rangle \otimes 1) \otimes 1)(x \otimes 1)$$

$$= (k(g \otimes 1)a^{-1})(\langle f\rangle \otimes 1)(a(x \otimes 1)),$$

which is again of type $\langle\ \rangle$, provided only that $k(g \otimes 1)a^{-1}$ is constructible. Since a^{-1} is central, we need the constructibility of the composite

$$(C \otimes D) \otimes U \xrightarrow[g \otimes 1]{} S \otimes U \xrightarrow[k]{} V\,;$$

this follows by the inductive hypothesis since T, whose rank is equal by Lemma 6.3 to that of $([B, C] \otimes A) \otimes D$, has been replaced by $C \otimes D$, clearly of lower rank.

The same induction shows that k and $g \otimes 1$ are compatible; so by Lemma 5.7 and Lemma 5.4, k is compatible with $(g \otimes 1)((\langle f \rangle \otimes 1) \otimes 1)(x \otimes 1) = h \otimes 1$.

Case 3: h is of type $\otimes$. Let h be $y(f \otimes g)x$ as in Proposition 6.2 (ii). We are to consider the composite $k(h \otimes 1)$; without loss of generality we may suppose that $x = 1$ and absorb $y \otimes 1$ into k. Then we have

$$\begin{aligned} k(h \otimes 1) &= k((f \otimes g) \otimes 1) = ka^{-1}(f \otimes (g \otimes 1))a \\ &= ka^{-1}(f \otimes 1)(1 \otimes (g \otimes 1))a\,, \end{aligned}$$

so that finally $k(h \otimes 1)$ is the composite

$$(6.1) \qquad (A \otimes B) \otimes U \xrightarrow[wa]{} B \otimes (A \otimes U) \xrightarrow[g \otimes 1]{} D \otimes (A \otimes U) \xrightarrow[ka^{-1}(f \otimes 1)w^{-1}]{} V\,,$$

where the w's stand for two instances of the central morphism aca^{-1}. Now the composite

$$A \otimes (D \otimes U) \xrightarrow[f \otimes 1]{} C \otimes (D \otimes U) \xrightarrow[ka^{-1}]{} V$$

is constructible by the induction hypothesis, because U has been replaced by $D \otimes U$ (*this* is the reason for formulating Proposition 6.4 for $k(h \otimes 1)$ instead of just kh) and we have

$$\begin{aligned} r(T) + r(S) + r(U) + r(V) &= r(A \otimes B) + r(C \otimes D) + r(U) + r(V) \\ &> r(A) + r(C) + r(D \otimes U) + r(V) \end{aligned}$$

unless $r(B) = 0$; in the latter case we get equality, but then $r(D) > 0$ since g is non-trivial, and

$$r(T) + r(S) = r(A \otimes B) + r(C \otimes D) > r(A) + r(C)\,,$$

so that the second half of the induction hypothesis applies. Since w^{-1} is central, $k' = ka^{-1}(f \otimes 1)w^{-1}$ is constructible; and since wa is central, (6.1) will be constructible if the composite $k'(g \otimes 1)$ is. But now the induction hypothesis shows that $k'(g \otimes 1)$ is indeed constructible, by essentially the same calculation with ranks as above, with g replacing f.

The same inductions show that ka^{-1} is compatible with $f \otimes 1$, so that k is compatible with $a^{-1}(f \otimes 1)w^{-1}$; and that k' is compatible with $g \otimes 1$, and hence with $(g \otimes 1)wa$. Therefore, by Lemma 5.3, k is compatible with $a^{-1}(f \otimes 1)w^{-1}(g \otimes 1)wa = h \otimes 1$.

Case 4: k is of type π. Thus $k = y\pi(f)$ for central y and constructible f. We can take $y = 1$, so that $k = \pi(f)\colon S \otimes U \to V = [B, C]$ for some constructible $f\colon (S \otimes U) \otimes B \to C$. By (5.2) we have

$$k(h \otimes 1) = \pi(f)(h \otimes 1) = \pi(f((h \otimes 1) \otimes 1))\,;$$

this is constructible if $f((h \otimes 1) \otimes 1)$ is, and hence if $f((h \otimes 1) \otimes 1)a^{-1} = fa^{-1}(h \otimes 1)$ is. This last is the composite

$$T \otimes (U \otimes B) \xrightarrow[h \otimes 1]{} S \otimes (U \otimes B) \xrightarrow[fa^{-1}]{} C\,,$$

which is constructible by induction, since $r(U \otimes B) + r(C) < r(U) + r([B, C])$.

The same induction proves fa^{-1} compatible with $h \otimes 1$, hence f with $a^{-1}(h \otimes 1)a = (h \otimes 1) \otimes 1$; and thence, by Lemma 5.6, $\pi(f)$ with $h \otimes 1$.

Case 5: h is of type π and k of type $\otimes$. There are central morphisms x, y and z such that $h = z\pi(m)$ for some constructible $m\colon T \otimes P \to Q$ and $k = y(f \otimes g)x$ for some constructible and non-trivial f and g. We may take $y = 1$ and absorb $z \otimes 1$ into x, so that the composite $k(h \otimes 1)$ to be considered has the form

$$T \otimes U \xrightarrow[h \otimes 1]{} [P, Q] \otimes U \xrightarrow[x]{} A \otimes B \xrightarrow[f \otimes g]{} C \otimes D = V\,.$$

Interchanging A and B if necessary, we can assume that the central morphism x associates $[P, Q]$ with a prime factor of A. Then Proposition 4.10 gives a shape R such that x has the form of a composite

$$[P, Q] \otimes U \xrightarrow[1 \otimes s]{} [P, Q] \otimes (R \otimes B) \xrightarrow[a^{-1}]{} ([P, Q] \otimes R) \otimes B \xrightarrow[t \otimes 1]{} A \otimes B\,,$$

for suitable central s and t. Since $(1 \otimes s)(h \otimes 1) = (h \otimes 1)(1 \otimes s)$ we can drop s and write $U = R \otimes B$, while t can be absorbed into f. The composite $k(h \otimes 1)$ to be considered now has the form

$$T \otimes (R \otimes B) \xrightarrow[h \otimes 1]{} [P, Q] \otimes (R \otimes B) \xrightarrow[a^{-1}]{} ([P, Q] \otimes R) \otimes B \xrightarrow[f \otimes g]{} C \otimes D .$$

This may be rewritten as

$$(f \otimes g) a^{-1} (h \otimes 1) = (f \otimes g)((h \otimes 1) \otimes 1) a^{-1} = (f(h \otimes 1) \otimes g) a^{-1} ,$$

which will be constructible by CM2 and CM3 if the composite

$$T \otimes R \xrightarrow[h \otimes 1]{} [P, Q] \otimes R \xrightarrow[f]{} C$$

is. That this is indeed so follows by induction, since g is non-trivial and therefore $r(U) + r(V) = r(R \otimes B) + r(C \otimes D) > r(R) + r(C)$.

The induction argument also shows that f is compatible with $h \otimes 1$, whence, by Lemmas 5.5 and 5.4, $f \otimes g$ is compatible with $((h \otimes 1) \otimes 1) a^{-1} = a^{-1}(h \otimes 1)$; finally, by Lemma 5.4 again, $k = (f \otimes g) a^{-1}$ is compatible with $h \otimes 1$.

Case 6: h is of type π and k is of type $\langle\ \rangle$. Thus there are central morphisms z and x such that $h = z\pi(m)$ and $k = g(\langle f \rangle \otimes 1)x$ for constructible morphisms

$$m \colon T \otimes P \to Q , \quad f \colon A \to B , \quad g \colon C \otimes D \to V .$$

By absorbing $z \otimes 1$ in x, we may suppose that $S = [P, Q]$ and that $z = 1$. Then the composite to be considered has the form

$$T \otimes U \xrightarrow[\pi(m) \otimes 1]{} [P, Q] \otimes U \xrightarrow[x]{} ([B, C] \otimes A) \otimes D \xrightarrow[g(\langle f \rangle \otimes 1)]{} V .$$

We distinguish three subcases, according as $[P, Q]$ is associated via the central morphism x with $[B, C]$, with a prime factor of A, or with a prime factor of D. (We recall that these possibilities need not be mutually exclusive, if $[P, Q]$ is constant.)

Subcase 1: $[P, Q]$ is associated with $[B, C]$. By Proposition 4.11, $P = B$, $Q = C$, and x is the composite

$$[B, C] \otimes U \xrightarrow[1 \otimes s]{} [B, C] \otimes (A \otimes D) \xrightarrow[a^{-1}]{} ([B, C] \otimes A) \otimes D$$

for a suitable central s. Since $(1 \otimes s)(h \otimes 1) = (h \otimes 1)(1 \otimes s)$ we may, arguing as in Case 5, suppose that $U = A \otimes D$ and $s = 1$. The composite to be considered then has the form

$$T\otimes(A\otimes D)\xrightarrow[\pi(m)\otimes 1]{}[B,C]\otimes(A\otimes D)\xrightarrow[a^{-1}]{}([B,C]\otimes A)\otimes D\xrightarrow[g(\langle f\rangle\otimes 1)]{}V.$$

Now $a^{-1}(\pi(m)\otimes 1)=((\pi(m)\otimes 1)\otimes 1)a^{-1}$ by naturality, while by Lemma 5.2 we have $\langle f\rangle(\pi(m)\otimes 1)=m(1\otimes f)$. The composite thus becomes

$$k(h\otimes 1)=g(m(1\otimes f)\otimes 1)a^{-1}=g(mc(f\otimes 1)c\otimes 1)a^{-1}\ .$$

This formula involves two successive composites, first a composite h',

$$A\otimes T\xrightarrow[f\otimes 1]{}B\otimes T\xrightarrow[mc]{}C=Q$$

and second the composite

$$(T\otimes A)\otimes D\xrightarrow[h'c\otimes 1]{}C\otimes D\xrightarrow[g]{}V\ .$$

Both are of the form considered in our induction. The induction assumption does apply to both because the original rank, with $U=A\otimes D$ and $S=[P,Q]=[B,C]$, is

$$rT+rS+rU+rV=rT+rV+rB+rC+rA+rD+1\ ,$$

and this clearly exceeds either of the ranks $rA+rB+rT+rC$ or $rT+rA+rD+rC+rV$ involved in the two composites above.

The same induction shows that g is compatible with $h'c\otimes 1$ and mc is compatible with $f\otimes 1$; so that by Lemmas 5.3, 5.4 and 5.5 $g(m\otimes 1)$ is compatible with $(1\otimes f)\otimes 1$. It follows from Lemma 5.9 and Lemma 5.4 that $g(\langle f\rangle\otimes 1)$ is compatible with $((\pi(m)\otimes 1)\otimes 1)a^{-1}=a^{-1}(\pi(m)\otimes 1)$, so that $k=g(\langle f\rangle\otimes 1)a^{-1}$ is compatible with $h\otimes 1=\pi(m)\otimes 1$, as required.

Subcase 2: $[P,Q]=S$ is associated with a prime factor of A. By Proposition 4.10, there is a shape R such that $a(c\otimes 1)x$ is the composite

$$S\otimes U\xrightarrow[1\otimes s]{}S\otimes(R\otimes([B,C]\otimes D))$$

$$\xrightarrow[a^{-1}]{}(S\otimes R)\otimes([B,C]\otimes D)\xrightarrow[t\otimes 1]{}A\otimes([B,C]\otimes D)$$

for suitable central s and t. By the naturality of a and c, therefore, x is the composite

$$S\otimes U\xrightarrow[1\otimes s]{}S\otimes(R\otimes([B,C]\otimes D))$$

$$\xrightarrow[w]{}([B,C]\otimes(S\otimes R))\otimes D\xrightarrow[(1\otimes t)\otimes 1]{}([B,C]\otimes A)\otimes D,$$

where w is the central natural transformation $w = (c \otimes 1)a^{-1}a^{-1}$. Once again, since $(1 \otimes s)(h \otimes 1) = (h \otimes 1)(1 \otimes s)$, we may suppose that $s = 1$ and $U = R \otimes ([B, C] \otimes D)$. Moreover, since $\langle f \rangle (1 \otimes t) = \langle ft \rangle$ by (5.5) (the naturality of $\langle\ \rangle$), we may absorb t in f and hence suppose that $A = S \otimes R$ and $t = 1$. The desired composite $k(h \otimes 1)$ thus has the form

$$\begin{aligned} g(\langle f \rangle \otimes 1)\, w(h \otimes 1) &= g(\langle f \rangle \otimes 1)((1 \otimes (h \otimes 1)) \otimes 1)\, w \\ &= g(\langle f(h \otimes 1) \rangle \otimes 1) w, \end{aligned}$$

by the naturality of w and of $\langle\ \rangle$. It thus suffices by CM5 to prove the composite

$$T \otimes R \xrightarrow[h \otimes 1]{} S \otimes R \xrightarrow[f]{} B$$

constructible. But this is of the form considered in the induction, and since $r(U) > r(R) + r(B)$ the inductive hypothesis applies.

By the same inductive argument f is compatible with $h \otimes 1$; by Lemmas 5.8 and 5.4, therefore, $g(\langle f \rangle \otimes 1)$ is compatible with $((1 \otimes (h \otimes 1)) \otimes 1)w = w(h \otimes 1)$; finally, by Lemma 5.4 again, $k = g(\langle f \rangle \otimes 1)\, w$ is compatible with $h \otimes 1$.

Subcase 3: $[P, Q] = S$ is associated with a prime factor of D. By Proposition 4.10, there is a shape R such that cx is the composite

$$S \otimes U \xrightarrow[1 \otimes s]{} S \otimes (R \otimes ([B, C] \otimes A))$$

$$\xrightarrow[a^{-1}]{} (S \otimes R) \otimes ([B, C] \otimes A) \xrightarrow[t \otimes 1]{} D \otimes ([B, C] \otimes A)$$

for suitable central s and t. By the naturality of c, therefore, x is the composite

$$S \otimes U \xrightarrow[1 \otimes s]{} S \otimes (R \otimes ([B, C] \otimes A))$$

$$\xrightarrow[u]{} ([B, C] \otimes A) \otimes (S \otimes R) \xrightarrow[1 \otimes t]{} ([B, C] \otimes A) \otimes D\,,$$

where u is the central natural transformation $u = ca^{-1}$. Since $(1 \otimes s)(h \otimes 1) = (h \otimes 1)(1 \otimes s)$ we may again suppose that $s = 1$, so that $U = R \otimes ([B, C] \otimes A)$. Since $(\langle f \rangle \otimes 1)(1 \otimes t) = (1 \otimes t)(\langle f \rangle \otimes 1)$, we may absorb $1 \otimes t$ in g and hence suppose that $t = 1$ and $D = S \otimes R$. The desired composite $k(h \otimes 1)$ is then

$$\begin{aligned} g(\langle f \rangle \otimes 1)\, u(h \otimes 1) &= g(\langle f \rangle \otimes 1)(1 \otimes (h \otimes 1))\, u \\ &= g(1 \otimes (h \otimes 1))(\langle f \rangle \otimes 1)\, u \\ &= gu(h \otimes 1)u^{-1}(\langle f \rangle \otimes 1)\, u\,, \end{aligned}$$

using (5.1) and the naturality of u. It thus suffices by CM5 to prove the constructibility of $gu(h \otimes 1)u^{-1}$, and therefore of $gu(h \otimes 1)$. This is the composite

$$T \otimes (R \otimes C) \xrightarrow[h \otimes 1]{} S \otimes (R \otimes C) \xrightarrow[gu]{} V,$$

which is constructible by the inductive hypothesis since $r(R \otimes C) < r(U)$.

By the same inductive argument gu is compatible with $h \otimes 1$, so that g is compatible with $u(h \otimes 1)u^{-1} = 1 \otimes (h \otimes 1)$. By Lemmas 5.8 and 5.4, therefore, $g(\langle f \rangle \otimes 1)$ is compatible with $(1 \otimes (h \otimes 1))u = u(h \otimes 1)$; so that finally $k = g(\langle f \rangle \otimes 1)u$ is compatible with $h \otimes 1$.

This concludes the proof of Proposition 6.4.

Theorem 6.5. *The constructible morphisms of $\underline{H}$ are exactly the allowable ones.*

Proof. We have already observed that the allowable morphisms clearly satisfy CM1–CM5, so that every constructible morphism is allowable. It remains to show that the constructible morphisms satisfy AM1–AM5.

They satisfy AM1 because $1, a, b, c, a^{-1}, b^{-1}$ are central. As for AM2, $d: T \to [S, T \otimes S]$ is $\pi(1)$ where $1: T \otimes S \to T \otimes S$, so that d is constructible by CM4; and $e: [T, S] \otimes T \to S$ is $\langle 1 \rangle$, which by the naturality of b is the composite

$$[T, S] \otimes T \xrightarrow[b^{-1}]{} ([T, S] \otimes T) \otimes I \xrightarrow[\langle 1 \rangle \otimes 1]{} S \otimes I \xrightarrow[b]{} S,$$

so that e is constructible by CM2 and CM5. AM3 is trivially satisfied, as it coincides with CM3. In AM4, let $f: T \to T'$ and $g: S \to S'$ be constructible. Then the composite

$$[T', S] \otimes T \xrightarrow[b^{-1}]{} ([T', S] \otimes T) \otimes I \xrightarrow[\langle f \rangle \otimes 1]{} S \otimes I \xrightarrow[gb]{} S'$$

is constructible by CM2 and CM5; but this composite is $g\langle f \rangle$ by the naturality of b. It follows from CM4 that $\pi(g\langle f \rangle)$ is constructible; but $\pi(g\langle f \rangle) = \pi(ge(1 \otimes f))$ is equal by the naturality of π to $[g, f]\ \pi(e) = [g, f]\, 1 = [g, f]$. Thus AM4 is satisfied.

There remains AM5. Let $f: T \to S$ and $g: S \to R$ be constructible. Then the composite

$$S \otimes I \xrightarrow[b]{} S \xrightarrow[g]{} R \tag{6.2}$$

is constructible by CM2, whence the composite

$$T \otimes I \xrightarrow[f \otimes 1]{} S \otimes I \xrightarrow[gb]{} R \tag{6.3}$$

is constructible by Proposition 6.4; by CM2 again, the composite of (6.3) with $b^{-1}: T \to T \otimes I$ is also constructible, and by the naturality of b this composite is gf.

Proof of Theorem 2.2. Let $f: T \to S$, $g: S \to R$ be allowable graphs in $\underline{G}$. By Theorem 6.5, they are constructible. The composite (6.2) is then also constructible, so that in (6.3) gb is compatible with $f \otimes 1$ by Proposition 6.4. We conclude from Lemma 5.4 that g is compatible with $b(f \otimes 1)b^{-1} = f$.

§7. Proofs of Theorem 2.1 and Theorem 2.4

We still use $\underline{H}$ to denote $\underline{N}(\underline{V})$ or $\underline{G}$, with $\Gamma: \underline{H} \to \underline{G}$ as before. For the purposes of this section we need a slight refinement of Proposition 6.2. Let us call an integral shape P *reduced* if it is either the shape I or else is constructed by the rules S2 and S3 alone; that is, it contains no I's unless it reduces to I alone. By *an iterated tensor product* of shapes $X_1, ..., X_n$ we mean $|P|(X_1, ..., X_n)$ for any reduced integral shape P with $v(P) = \{1, ..., n\}$; if $n = 0$ it is just I. Let us call an arbitrary shape T *reduced* if, in its prime factorization $T = |P|(X_1, ..., X_n)$, the integral shape P is reduced. (This is a consistent use of language since the prime factorization of the integral shape P is $|P|(1, 1, ..., 1)$.)

Lemma 7.1. *Given any shape T we can find a reduced shape T' and a central isomorphism $z: T \to T'$ in $\underline{H}$.*

Proof. Let the prime factorization of T be $|P|(X_1, ..., X_n)$. Let P' be a reduced integral shape with $v(P') = v(P) = \{1, ..., n\}$, and let $\xi: P \to P'$ be the graph corresponding to the identity permutation of $\{1, ..., n\}$. Set $T' = |P'|(X_1, ..., X_n)$ and set $z = |\xi|_{\underline{H}}(X_1, ..., X_n)$.

Lemma 7.2. *In Proposition 6.2 we can suppose that the shapes A, B, C, D in (ii) and the shapes A, D in (iv) are reduced.*

Proof. In case (ii), replace A, B, C, D by reduced isomorphs as in Lemma 7.1, absorbing the central isomorphims thereby introduced into x and y; similarly for case (iv).

We define the *rank* $r(h)$ of a morphism $h: T \to S$ in $\underline{H}$ to be the sum $r(T) + r(S)$ of the ranks of T and of S. If h is allowable, which by Theorem 6.5 is the same thing as constructible, Proposition 6.2 asserts that h has one of the four following forms:

$$(7.1) \qquad h = x\,, \quad h = y(f \otimes g)x\,, \quad h = y\pi(f)\,, \quad h = g(\langle f \rangle \otimes 1)x\,,$$

where x and y are central, f and g are allowable, and moreover in the $y(f \otimes g)x$ case neither f nor g is trivial. The basis of our inductive arguments is the obvious fact that in each case we have $r(f) < r(h)$ and (where applicable) $r(g) < r(h)$. In using the forms (7.1) we shall always suppose that the reductions of Lemma 7.2 have been carried out.

Proof of Theorem 2.1. We are to construct an algorithm for deciding whether a graph $h: T \to S$ in $\underline{G}$ is allowable. We suppose inductively that we possess such an algorithm for all smaller values, if any, of $r(h)$. Since finding the prime factorizations of T and of S is algorithmic, Propositions 4.3 and 4.2 enable us to decide whether h is central. It remains to test whether h is of one of the remaining types

in (7.1), which we again refer to as type $\otimes$, type π, and type $\langle\ \rangle$.

To test whether h is of type $\otimes$, with the notation as in Proposition 6.2 and with A, B, C, D reduced, first observe that, by Proposition 4.3, the prime factors of A and of B must together make up those of T; so that there are only a finite number of possibilities for A and for B to be tried (because A and B are reduced!). Similarly there are only a finite number of possibilities for C and for D; and for a given choice of A, B, C, D there are only a finite number of possibilities for x and for y. When these choices are all made, the graph $y^{-1}hx^{-1}$ is either not of the form $f \otimes g$, or else is of this form for a unique f and g. Since $r(f) < r(h)$ and $r(g) < r(h)$, we can now test f and g for allowability.

Entirely similar procedures allow us to test whether h is of type π or of type $\langle\ \rangle$, so that we have the desired algorithm.

Remark. There are non-allowable graphs in $\underline{G}$; the unique graph $[1, 1] \to I$ is one such.

Before proving Theorem 2.4 we establish some facts about proper shapes, as defined in §2. Observe that every constant shape is proper; that if $[T, S]$ is proper then T and S are proper; and that $T \otimes S$ is proper if and only if T and S are proper – whence T is proper if and only if each of its prime factors is proper.

Lemma 7.3. *If $h: T \to S$ is a central morphism in $\underline{H}$ and if either T or S is proper so is the other.*

Proof. By Proposition 4.3, T and S have the same prime factors.

Proposition 7.4. *Let $h: T \to S$ be allowable in $\underline{H}$, with the shape S constant and the shape T proper. Then the shape T is constant.*

Proof. Suppose inductively that it is so for all smaller values, if any, of $r(h)$, and consider h of one of the four possible types in (7.1). If h is central the result is immediate by Corollary 4.4. By this same Corollary 4.4, together with Lemma 7.3 and Lemma 6.3, we may ignore central factors x and y in the other types in (7.1). If $h = f \otimes g: A \otimes B \to C \otimes D$ then C and D are constant because S is and A and B are proper because T is, so that by induction A and B are constant, whence T is constant. If $h = \pi(f): T \to [B, C]$ then B and C are constant because S is, so that $T \otimes B$ is proper because T is, and then T is constant by the inductive hypothesis applied to $f: T \otimes B \to C$. Finally, if $h = g(\langle f \rangle \otimes 1): ([B, C] \otimes A) \otimes D \to S$, then the inductive hypothesis applied to $g: C \otimes D \to S$ shows that C and D are constant. Since T is proper so is $[B, C]$, whence B is constant. Finally the inductive hypothesis applied to $f: A \to B$ shows that A is constant, so that T is constant.

We next show how to eliminate constant prime factors from a shape T.

Lemma 7.5. *Given a shape T we can find a shape S with $r(S) \leq r(T)$ and an allowable isomorphism $f: T \to S$ in $\underline{H}$ with allowable inverse such that*

(a) *S is reduced;*

(b) *S has no constant prime factors, its prime factors being precisely the non-constant ones of T;*

(c) *if T is proper, so is S;*

(d) *there is a constant shape R and a central isomorphism $T \to S \otimes R$ with the same graph as f.*

Proof. Let S be any iterated tensor product of the non-constant prime factors of T, and R any iterated tensor product of the constant prime factors of T. There is an evident central isomorphism $T \to S \otimes R$, and f is the composite of this with

$$S \otimes R \xrightarrow[1 \otimes k_R]{} S \otimes I \xrightarrow[b]{} S,$$

where k_R is the isomorphism of Lemma 4.7.

Proposition 7.6. *Let $h: P \otimes Q \to M \otimes N$ be an allowable morphism in $\underline{H}$, where P, Q, M, N are proper shapes. Suppose that the graph Γh is of the form $\xi \otimes \eta$ for graphs $\xi: P \to M$ and $\eta: Q \to N$. Then there are allowable morphisms $p: P \to M$, $q: Q \to N$ such that $h = p \otimes q$, $\Gamma p = \xi$, and $\Gamma q = \eta$.*

Proof. Suppose inductively that it is so for all smaller values, if any, of $r(h)$. By Lemma 7.5 we may without loss of generality suppose each of P, Q, M, N to be reduced and to have only non-constant prime factors.

If h is central, so is $\Gamma h = \xi \otimes \eta$. From Propositions 4.3 and 4.2, it is clear that ξ and η are then central. By Theorem 4.9 there are central $p: P \to M$, $q: Q \to N$ with $\Gamma p = \xi$ and $\Gamma q = \eta$; then $\Gamma h = \Gamma(p \otimes q)$, so that by Theorem 4.9 again we have $h = p \otimes q$.

If h is of type $\otimes$, say h is the composite

$$P \otimes Q \xrightarrow[x]{} A \otimes B \xrightarrow[f \otimes g]{} C \otimes D \xrightarrow[y]{} M \otimes N;$$

let an iterated tensor product, in the order in which they occur in P, of those prime factors of P that are associated via x with a prime factor of A [resp. B] be X [resp. Y]; similarly let an iterated $\otimes$-product of those prime factors of Q associated via x with a prime factor of A [resp. B] be U [resp. V]. In the same way let X', Y', U', V' be iterated $\otimes$-products of the prime factors "common" to M and C, M and D, N and C, N and D respectively. Define a graph $\rho: X \to X'$ as the restriction of Γh to $v(X) + v(X')$; this is indeed a graph because Γh is of the form $\xi \otimes \eta$. Define similarly graphs $\sigma: Y \to Y'$, $\tau: U \to U'$, $\kappa: V \to V'$. The graphs of the allowable morphisms

(7.2) $$X \otimes U \longrightarrow A \xrightarrow[f]{} C \longrightarrow X' \otimes U' ,$$

(7.3) $$Y \otimes V \longrightarrow B \xrightarrow[g]{} D \longrightarrow Y' \otimes V' ,$$

where the unnamed arrows denote the obvious central morphisms, are respectively $\rho \otimes \tau$ and $\sigma \otimes \kappa$. By the inductive hypothesis we conclude that (7.2) and (7.3) are respectively $r \otimes t$ and $s \otimes k$ for allowable morphisms r, t, s, k with the respective graphs $\rho, \tau, \sigma, \kappa$. Define p and q to be the composites

$$P \longrightarrow X \otimes Y \xrightarrow[r \otimes s]{} X' \otimes Y' \longrightarrow M ,$$

$$Q \longrightarrow U \otimes V \xrightarrow[t \otimes k]{} U' \otimes V' \longrightarrow N ,$$

where once again the unnamed arrows denote the obvious central morphisms. That $h = p \otimes q$ is then immediate from Theorem 4.9, while evidently $\Gamma p = \xi$ and $\Gamma q = \eta$.

If h is of type π, say h is the composite

$$P \otimes Q \xrightarrow[\pi(f)]{} [B, C] \xrightarrow[y]{} M \otimes N ,$$

then by Proposition 4.3 either $M = [B, C]$ and $N = I$ or else $N = [B, C]$ and $M = I$; by replacing h by chc if necessary we may suppose the former to be the case. Then h is the composite

$$P \otimes Q \xrightarrow[\pi(f)]{} [B, C] \xrightarrow[b^{-1}]{} [B, C] \otimes I .$$

Since $\Gamma h = \xi \otimes \eta$ it follows from Lemma 5.1 that the graph of the composite

$$(P \otimes B) \otimes Q \xrightarrow[u]{} (P \otimes Q) \otimes B \xrightarrow[f]{} C \xrightarrow[b^{-1}]{} C \otimes I ,$$

where u is the evident central morphism, is $\pi^{-1}(\xi) \otimes \eta$. By induction, therefore, $b^{-1} fu$ is $r \otimes q$ for allowable $r \colon P \otimes B \to C$ with graph $\pi^{-1}(\xi)$ and $q : Q \to I$ with graph η. Set $p = \pi(r)$; then $\Gamma p = \xi$ and $h = p \otimes q$ by another application of Lemma 5.1.

If h is of type $\langle\ \rangle$, with the notation of Proposition 6.2, we may (replacing h by chc if necessary) suppose that $[B, C]$ is associated via x with a prime factor of P. Let an iterated $\otimes$-product of those prime factors of A associated via x with a prime factor of P [resp. Q] be X [resp. Y]. The mate under Γh of an element of $v(Y)$ is in $v(A) + v(B)$ by the form $g(\langle f \rangle \otimes 1)x$ of h, but is in $v(Q) + v(N)$ by the hypothesis that $\Gamma h = \xi \otimes \eta$; it must therefore be in $v(Y)$. Thus the graph of the composite

(7.4) $$X \otimes Y \longrightarrow A \xrightarrow[f]{} B \xrightarrow[b^{-1}]{} B \otimes I ,$$

where the unnamed arrow is the obvious central morphism, is of the form $\rho \otimes \sigma$ for graphs $\rho : X \to B$ and $\sigma : Y \to I$. By the inductive hypothesis, (7.4) is $r \otimes s$ for allowable $r : X \to B$, $s : Y \to I$. It then follows from Proposition 7.4 that Y is constant; since none of the prime factors of Q is constant, this means that $Y = I$ and that all the prime factors of A are therefore associated via x with prime factors of P.

It follows then from Proposition 4.10 that there are a shape R and central morphisms y and z such that x is the composite

$$P \otimes Q \xrightarrow[y \otimes 1]{} (([B,C] \otimes A) \otimes R) \otimes Q \xrightarrow[a]{} ([B,C] \otimes A) \otimes (R \otimes Q)$$
$$\xrightarrow[1 \otimes z]{} ([B,C] \otimes A) \otimes D .$$

It is clear that the graph of the composite

$$(C \otimes R) \otimes Q \xrightarrow[a]{} C \otimes (R \otimes Q) \xrightarrow[1 \otimes z]{} C \otimes D \xrightarrow[g]{} M \otimes N$$

is $\zeta \otimes \eta$, where $\zeta : C \otimes R \to M$ is the restriction of ξ to $v(C) + v(R) + v(M)$. By induction, $g(1 \otimes z)a$ is $r \otimes q$ for allowable $r : C \otimes R \to M$ and $q : Q \to N$ with the appropriate graphs. Setting p equal to the composite

$$P \xrightarrow[y]{} ([B,C] \otimes A) \otimes R \xrightarrow[r(\langle f \rangle \otimes 1)]{} M ,$$

we have $p \otimes q = (r \otimes q)((\langle f \rangle \otimes 1) \otimes 1)(y \otimes 1) = g(1 \otimes z)a(((\langle f \rangle \otimes 1) \otimes 1)(y \otimes 1)$; by the naturality of a this is $g(1 \otimes z)(\langle f \rangle \otimes 1)a(y \otimes 1) = g(\langle f \rangle \otimes 1)(1 \otimes z)a(y \otimes 1) = g(\langle f \rangle \otimes 1)x = h$.

This completes the proof of Proposition 7.6.

Proposition 7.7. *Let $f : A \otimes B \to C$ be an allowable morphism in $\underline{H}$, where A, B, C are proper shapes. Suppose that the mate under Γf of each element of $v(B)$ is again in $v(B)$. Then B is constant.*

Proof. The composite

$$A \otimes B \xrightarrow[f]{} C \xrightarrow[b^{-1}]{} C \otimes I$$

has a graph of the form $\xi \otimes \eta$; therefore by Proposition 7.6 there is an allowable morphism $q : B \to I$. It follows from Proposition 7.4 that B is constant.

Proposition 7.8. *Let $h : ([Q,M] \otimes P) \otimes N \to S$ be an allowable morphism between proper shapes in $\underline{H}$, with $[Q,M]$ not constant. Suppose that the graph Γh is of the form $\eta(\langle \xi \rangle \otimes 1)$ for graphs $\xi : P \to Q$, $\eta : M \otimes N \to S$. Suppose finally that ξ cannot be written in the form*

$$(7.5) \qquad P \xrightarrow[\omega]{} ([F,G] \otimes E) \otimes H \xrightarrow[\rho(\langle \omega \rangle \otimes 1)]{} Q$$

for any graphs ω, ρ, σ with ω central. Then there are allowable morphisms $p: P \to Q$, $q: M \otimes N \to S$ such that $h = q(\langle p \rangle \otimes 1)$, $\Gamma p = \xi$ and $\Gamma q = \eta$.

Proof. Suppose inductively that it is so for all smaller values, if any, of $r(h)$. Use Lemma 7.5 to replace P, N, S by reduced shapes which have no constant prime factors; we must show that in doing so we lose no generality. It follows from (5.5) that doing so makes no difference to the expressibility Γh in the form $\eta(\langle \xi \rangle \otimes 1)$ or the expressibility of h in the form $q(\langle p \rangle \otimes 1)$ for allowable p and q. We must show that it makes no difference to the expressibility of ξ in the form (7.5); but this is a very easy deduction from Lemma 7.5 (d).

Note that once we have $h = q(\langle p \rangle \otimes 1)$, it is automatic that $\Gamma p = \xi$ and $\Gamma q = \eta$.

Suppose that h is central. By Corollary 4.5, the mate under Γh of an element of $v(P)$ or of $v(Q)$ is then an element of $v(S)$. On the other hand, by the form $\eta(\langle \xi \rangle \otimes 1)$ of Γh the mate of an element of $v(P)$ is an element of $v(Q)$, and conversely. It follows that P and Q are both constant, so that by the reduction above we must have $P = I$. Lemma 4.7 now gives an allowable $p = k_Q^{-1} : I \to Q$. By the naturality of e,

$$\langle p \rangle = e(1 \otimes p) = e([p, 1] \otimes 1) : [Q, M] \otimes I \longrightarrow [I, M] \otimes I \longrightarrow M \,.$$

On the other hand, $[A, [I, M]] \cong [A \otimes I, M] \cong [A, M]$ in any closed category, so the Yoneda Lemma provides an isomorphism $[1, b]d : M \cong [I, M]$ with inverse eb^{-1}. Therefore $[1, b]de = b : [I, M] \otimes I \cong [I, M]$, so that from the display above

$$[1, b]d\langle p \rangle = b([p, 1] \otimes 1) : [Q, M] \otimes I \longrightarrow [I, M] \,.$$

Since the right-hand morphism is an isomorphism we have constructed the following factorization of the identity

$$1 = t\langle p \rangle : [Q, M] \otimes I \longrightarrow [Q, M] \otimes I \,,$$

with an allowable $t = ([p^{-1}, 1] \otimes 1) b^{-1} [1, b] d$. The originally given allowable morphism h can now be factored as

$$h = h1 = h(t \otimes 1)(\langle p \rangle \otimes 1) = q(\langle p \rangle \otimes 1)$$

with q allowable, as required.

If h is of the form $y(f \otimes g)x$ for $f: A \to C$, $g: B \to D$, we may without loss of generality suppose that $[Q, M]$ is associated via x with a prime factor of A. Let an iterated $\otimes$-product of those prime factors of P associated via x with a prime factor of A [resp. B] be X [resp. Y]. The mate under Γh of an element of $v(Y)$ is in $v(B) + v(D)$ by the form $y(f \otimes g)x$ of h, but is in $v(Q) + v(P)$ by the hypothesis that $\Gamma h = \eta(\langle \xi \rangle \otimes 1)$; it must therefore be in $v(Y)$. It now follows from Proposition

7.7 that Y is constant; since none of the prime factors of P is constant, this means that $Y=I$ and that all the prime factors of P are associated via x with prime factors of A.

It follows then from Proposition 4.10 that there are a shape R and central morphisms s and t such that x is the composite

$$([Q,M]\otimes P)\otimes N \xrightarrow[1\otimes t]{} ([Q,M]\otimes P)\otimes(R\otimes B)$$
$$\xrightarrow[a^{-1}]{} (([Q,M]\otimes P)\otimes R)\otimes B \xrightarrow[s\otimes 1]{} A\otimes B\,.$$

It is clear that the graph of the composite

$$([Q,M]\otimes P)\otimes R \xrightarrow[s]{} A \xrightarrow[f]{} C$$

is $\zeta(\langle\xi\rangle\otimes 1)$, where $\zeta: M\otimes R\to C$ is the restriction of η to $\upsilon(M)+\upsilon(R)+\upsilon(C)$. It follows by induction that fs is $r(\langle p\rangle\otimes 1)$ for allowable $p: P\to Q$ and $r: M\otimes R\to C$. Then

$$h = y(f\otimes g)x = y(f\otimes g)(s\otimes 1)a^{-1}(1\otimes t)$$
$$= y(fs\otimes g)a^{-1}(1\otimes t) = y(r\otimes g)((\langle p\rangle\otimes 1)\otimes 1)a^{-1}(1\otimes t)$$
$$= y(r\otimes g)a^{-1}(\langle p\rangle\otimes 1)(1\otimes t) = y(r\otimes g)a^{-1}(1\otimes t)(\langle p\rangle\otimes 1)$$

is of the required form, with $q = y(r\otimes g)a^{-1}(1\otimes t)$.

If h is of the form $y\pi(f)$ for some $y: [B,C]\to S$, we must have $[B,C]=S$ and $y=1$. Then $\Gamma h = \eta(\langle\xi\rangle\otimes 1)$, $\pi^{-1}h = f$ and the naturality of π^{-1} shows that $\Gamma f = \pi^{-1}\eta((\langle\xi\rangle\otimes 1)\otimes 1)$. If we rewrite this as

$$\Gamma(fa^{-1}) = (\Gamma f)a^{-1} = (\pi^{-1}\eta)a^{-1}(\langle\xi\rangle\otimes 1): ([Q,M]\otimes P)\otimes(N\otimes B)\to C\,,$$

we can apply the induction assumption to fa^{-1} to get $fa^{-1} = r(\langle p\rangle\otimes 1)$ for allowable $p: P\to Q$ and $r: M\otimes(N\otimes B)\to C$. Then $f = ra\ \ ((\langle p\rangle\otimes 1)\otimes 1)$, so by the naturality of π we get

$$h = \pi f = \pi(ra^{-1})(\langle p\rangle\otimes 1)\,,$$

which is in the desired form.

In the final case where h is of the form

$$([Q,M]\otimes P)\otimes N \xrightarrow[x]{} ([B,C]\otimes A)\otimes D \xrightarrow[\langle f\rangle\otimes 1]{} C\otimes D \xrightarrow[g]{} S\,,$$

we distinguish cases according as $[Q,M]$ is associated via x with (i) $[B,C]$; (ii) a prime factor of A; or (iii) a prime factor of D.

Case (i). By Proposition 4.11, $B = Q$, $C = M$, and x is the composite

$$([Q,M] \otimes P) \otimes N \xrightarrow[a]{} [Q,M] \otimes (P \otimes N)$$

$$\xrightarrow[1 \otimes y]{} [Q,M] \otimes (A \otimes D) \xrightarrow[a^{-1}]{} ([Q,M] \otimes A) \otimes D$$

for some central y. Let X be an iterated $\otimes$-product of those prime factors of P associated via y with a prime factor of D. The mate under Γh of an element of $v(X)$ is in $v(M) + v(D) + v(S)$ by the form $g(\langle f \rangle \otimes 1)x$ of h, but is in $v(P) + v(Q)$ by the hypothesis that $\Gamma h = \eta(\langle \xi \rangle \otimes 1)$; it must therefore be in $v(X)$. Then X is constant by Proposition 7.7, and since P has no constant prime factors, this means that all the prime factors of P are associated via y with prime factors of A. A similar argument shows that all the prime factors of N are associated via y with prime factors of D. We may therefore, absorbing central morphisms into g and f where necessary, suppose without loss of generality that $A = P$, $D = Q$, and $x = 1$. Then $h = q(\langle p \rangle \otimes 1)$ with $q = g$ and $p = f$.

Case (ii). $[Q,M]$ is associated via x with a prime factor of A. Suppose if possible that $[B,C]$ were associated via x with a prime factor of P. Let X be an iterated $\otimes$-product of all those prime factors of $([Q,M] \otimes P) \otimes N$ that either are prime factors of P or else are associated via x with prime factors of A. The mate under Γh of an element of $v(A)$ is in $v(A) + v(B)$ by the form $g(\langle f \rangle \otimes 1)x$ of h, while the mate under Γh of an element of $v(P)$ is in $v(P) + v(Q)$ by the hypothesis that $\Gamma h = \eta(\langle \xi \rangle \otimes 1)$; thus the mate under Γh of an element of $v(X)$ is again in $v(X)$. It follows from Proposition 7.7 that X is constant, which contradicts the hypothesis that $[Q,M]$ is not constant.

Thus no prime factor of P is associated via x with $[B,C]$. Let Y be an iterated $\otimes$-product of those prime factors of P associated via x with prime factors of D. The mate under Γh of an element of $v(Y)$ is in $v(C) + v(D) + v(S)$ by the form $g(\langle f \rangle \otimes 1)x$ of h, but is in $v(P) + v(Q)$ by the hypothesis that $\Gamma h = \eta(\langle \xi \rangle \otimes 1)$; it must therefore be in $v(Y)$. Then Y is constant by Proposition 7.7, and since P has no constant prime factors this means that every prime factor of P is associated via x with a prime factor of A.

Then by Proposition 4.10 there are a shape R and central morphisms t and s such that $a(c \otimes 1)x$ is the composite

$$([Q,M] \otimes P) \otimes N \xrightarrow[1 \otimes s]{} ([Q,M] \otimes P) \otimes (R \otimes ([B,C] \otimes D))$$

$$\xrightarrow[a^{-1}]{} (([Q,M] \otimes P) \otimes R) \otimes ([B,C] \otimes D)$$

$$\xrightarrow[t \otimes 1]{} A \otimes ([B,C] \otimes D);$$

thus by naturality x is the composite

$$\begin{aligned}([Q,M]\otimes P)\otimes N &\xrightarrow[1\otimes n]{} ([Q,M]\otimes P)\otimes(([B,C]\otimes R)\otimes D)\\ &\xrightarrow[v]{} ([B,C]\otimes(([Q,M]\otimes P)\otimes R))\otimes D\\ &\xrightarrow[(1\otimes t)\otimes 1]{} ([B,C]\otimes A)\otimes D\,,\end{aligned}$$

where n is the central morphism $(c\otimes 1)a^{-1}s$ and v is the evident central morphism. It is clear that the graph of the composite

$$([Q,M]\otimes P)\otimes R \xrightarrow[t]{} A \xrightarrow[f]{} B$$

is $\zeta(\langle\xi\rangle\otimes 1)$, where $\zeta: M\otimes R\to B$ is the restriction of η to $v(M)+v(R)+v(B)$. So by induction ft is $r(\langle p\rangle\otimes 1)$ for allowable $p: P\to Q$ and $r: M\otimes R\to B$. Setting q equal to the composite

$$\begin{aligned}M\otimes N &\xrightarrow[1\otimes n]{} M\otimes(([B,C]\otimes R)\otimes D)\\ &\xrightarrow[v]{} ([B,C]\otimes(M\otimes R))\otimes D \xrightarrow[g(\langle r\rangle\otimes 1)]{} S\,,\end{aligned}$$

we have $q(\langle p\rangle\otimes 1) = g(\langle r\rangle\otimes 1)v(1\otimes n)(\langle p\rangle\otimes 1) = g(\langle r\rangle\otimes 1)\,v(\langle p\rangle\otimes 1)(1\otimes n)$; by the naturality of v this is $g(\langle r\rangle\otimes 1)((1\otimes(\langle p\rangle\otimes 1))\otimes 1)\,v(1\otimes n)$. Using (5.5), $\langle r\rangle(1\otimes(\langle p\rangle\otimes 1)) = \langle r(\langle p\rangle\otimes 1)\rangle$; which is $\langle ft\rangle$, or $\langle f\rangle(1\otimes t)$ by (5.5) again. Thus finally,

$$q(\langle p\rangle\otimes 1) = g(\langle f\rangle\otimes 1)((1\otimes t)\otimes 1)\,v(1\otimes n) = g(\langle f\rangle\otimes 1)x = h\,.$$

Case (iii). $[Q,M]$ is associated via x with a prime factor of D. Suppose if possible that $[B,C]$ were associated via x with a prime factor of P. Let X be an iterated $\otimes$-product of those prime factors of A that are associated via x with prime factors of N. The mate under Γh of an element of $v(X)$ is in $v(A)+v(B)$ by the form $g(\langle f\rangle\otimes 1)x$ of h, but is in $v(M)+v(N)+v(S)$ by the hypothesis that $\Gamma h = \eta(\langle\xi\rangle\otimes 1)$; it must therefore be in $v(X)$. Then X is constant by Proposition 7.7, and since N has no constant prime factors we conclude that every prime factor of A is associated via x with a prime factor of P. This implies that ξ is of the form

$$P \xrightarrow[\omega]{} ([B,C]\otimes A)\otimes H \xrightarrow[\rho(\langle\sigma\rangle\otimes 1)]{} Q$$

for some integral ω, which is excluded by hypothesis.

Thus no prime factor of P is associated via x with $[B,C]$. Let Y be an iterated $\otimes$-product of those prime factors of P associated via x with prime factors of A. The mate under Γh of an element of $v(Y)$ is in $v(A)+v(B)$ by the form $g(\langle f\rangle\otimes 1)x$ of h,

but is in $v(P) + v(Q)$ by the hypothesis that $\Gamma h = \eta(\langle\xi\rangle \otimes 1)$; it must therefore be in $v(Y)$. It follows from Proposition 7.7 that Y is constant, and since P has no constant prime factors this means that all the prime factors of P are associated via x with prime factors of D.

For brevity, let us write

$$M' = [Q, M] \otimes P, \qquad C' = [B, C] \otimes A$$

so that $\langle f \rangle : C' \to C$. In the central morphism $x : M' \otimes N \to C' \otimes D$, we now know that all the prime factors of M' are associated with prime factors of D. Apply Proposition 4.10 to this situation; it gives a shape R, a central morphism $N \to C' \otimes R$ (which without loss we can take to be the identity) and a central morphism $t : M' \otimes R \to D$, so that $x = (1 \otimes t)w$, as in the first row of the following diagram, in which w is the evident central morphism:

$$\begin{tikzcd}
M' \otimes (C' \otimes R) \arrow[r, "w"'] \arrow[d, "1 \otimes (\langle f \rangle \otimes 1)"] & C' \otimes (M' \otimes R) \arrow[r, "1 \otimes t"'] \arrow[d, "\langle f \rangle \otimes 1"] & C' \otimes D \arrow[d, "\langle f \rangle \otimes 1"] & \\
M' \otimes (C \otimes R) \arrow[r, "w"'] & C \otimes (M' \otimes R) \arrow[r, "1 \otimes t"'] & C \otimes D \arrow[r, "g"'] & S .
\end{tikzcd}$$

The diagram evidently commutes. Now the hypothesis $\Gamma h = \eta(\langle\xi\rangle \otimes 1)$ for $\eta : M \otimes N \to S$ clearly means that the graph of the composite $g(1 \otimes t)w$ is $\zeta(\langle\xi\rangle \otimes 1)$, where $\zeta : M \otimes (C \otimes R) \to S$ is the restriction of η to $vM + vC + vR + vS$. By induction, $g(1 \otimes t)w$ is $r(\langle p \rangle \otimes 1)$ for allowable $p : P \to Q$ and $r : M \otimes (C \otimes R) \to S$. Therefore, since $\langle p \rangle : M' \to M$,

$$h = g(1 \otimes t)w(1 \otimes (\langle f \rangle \otimes 1)) = r(\langle p \rangle \otimes 1)(1 \otimes (\langle f \rangle \otimes 1)) \quad .$$

$$= r(1 \otimes (\langle f \rangle \otimes 1))(\langle p \rangle \otimes 1)$$

has the requisite form $q(\langle p \rangle \otimes 1)$ for $q = r(1 \otimes (\langle f \rangle \otimes 1))$.

This concludes the proof of Proposition 7.8.

Proof of Theorem 2.4. Let T, S be proper shapes and let $h, h' : T \to S$ be allowable natural transformations in $\underline{N}(\underline{V})$ with $\Gamma h = \Gamma h'$; we are to prove that $h = h'$. Suppose inductively that it is so for all smaller values (if any) of $r(h)$; note that $r(h) = r(T) + r(S) = r(h')$. By Lemma 7.5, we may suppose that none of the prime factors of T or of S is constant.

If both h and h' are central we have $h = h'$ by Theorem 4.9. So we may suppose that h is of one of the other forms (7.1).

If h is of the form $y\pi(f)$, we have $y^{-1}h = \pi(f)$ where f is allowable. But then $y^{-1}h' = \pi(f')$, where $f' = \pi^{-1}(y^{-1}h')$ is also allowable by (1.2). Since h and h' have the same graph, so do f and f'; hence $f' = f$ by the inductive hypothesis, whence $h' = h$.

If h is of the form $y(f \otimes g)x$, we have $y^{-1}hx^{-1} = f \otimes g \colon A \otimes B \to C \otimes D$. Then $\Gamma(y^{-1}h'x^{-1}) = \Gamma(y^{-1}hx^{-1}) = \Gamma f \otimes \Gamma g$; so that by Proposition 7.6 $y^{-1}h'x^{-1} = f' \otimes g'$ for allowable $f' \colon A \to C$ and $g' \colon B \to D$ with $\Gamma f = \Gamma f'$ and $\Gamma g = \Gamma g'$, whence $f' = f$ and $g' = g$ by the inductive hypothesis. Hence $h' = h$.

There remains the case where h is of the form $g(\langle f \rangle \otimes 1)x$. Then it may be the case that the graph Γf of f is of the form

$$A \xrightarrow[\omega]{} ([F, G] \otimes E) \otimes H \xrightarrow[\rho(\langle \sigma \rangle \otimes 1)]{} B$$

for some central ω and some ρ, σ; in this case we have $\Gamma h = \Gamma g(\langle \Gamma f \rangle \otimes 1)\Gamma x$. Here, since $\langle\ \rangle$ is natural, as in (5.5),

$$\langle \Gamma f \rangle = \langle \rho(\langle \sigma \rangle \otimes 1)\omega \rangle = \langle \rho \rangle (1 \otimes (\langle \sigma \rangle \otimes 1))(1 \otimes \omega)\,.$$

It now follows easily that $\Gamma h = \tau(\langle \sigma \rangle \otimes 1)\psi$ for some τ and for some central ψ. Perhaps σ is of the form

$$E \xrightarrow[\phi]{} ([X, Y] \otimes Z) \otimes W \xrightarrow[\kappa(\langle \lambda \rangle \otimes 1)]{} F$$

for some central ϕ and some κ, λ; but E has strictly fewer prime factors than A, since $[F, G]$ is a prime factor of A but not of E; Z has strictly fewer prime factors than E; and so on. Thus this process terminates, and ultimately we have an expression for Γh of the form

$$T \xrightarrow[\mu]{} ([Q, M] \otimes P) \otimes N \xrightarrow[\eta(\langle \xi \rangle \otimes 1)]{} S$$

where μ is central and ξ is not of the form (7.5). Moreover $[Q, M]$ is not constant since T has no constant prime factors. By Theorem 4.9 there is a central natural transformation $y \colon T \to ([Q, M] \otimes P) \otimes N$ with $\Gamma y = \mu$. From Proposition 7.8 applied to hy^{-1} and $h'y^{-1}$ we conclude that $hy^{-1} = q(\langle p \rangle \otimes 1)$ and $h'y^{-1} = q'(\langle p' \rangle \otimes 1)$ for allowable $p, p' \colon P \to Q$ and allowable $q, q' \colon M \otimes N \to S$ with $\Gamma p = \Gamma p'$ and $\Gamma q = \Gamma q'$. It follows from the inductive hypothesis that $p = p'$ and $q = q'$, so that $h = h'$.

This completes the proof of Theorem 2.4.

References

[1] S.Eilenberg and G.M.Kelly, A generalization of the functorial calculus, J. Algebra 3 (1966) 366–375.
[2] S.Eilenberg and G.M.Kelly, Closed categories, in: Proc. Conf. on Categorical Algebra, La Jolla, 1965 (Springer-Verlag, 1966) pp. 421–562.
[3] D.B.A.Epstein, Functors between tensored categories, Invent. Math. 1 (1966) 221–228.
[4] G.Gentzen, Untersuchungen über das logische Schliessen I, II, Math. Z. 39 (1934–1935) 176–210 and 405–431.
[5] G.M.Kelly, On MacLane's conditions for coherence of natural associativities, commutativities, etc., J. Algebra 4 (1964) 397–402.
[6] J.Lambek, Deductive systems and categories I. Syntactic calculus and residuated categories, Math. Systems Theory 2 (1968) 287–318.
[7] J.Lambek, Deductive systems and categories II. Standard constructions and closed categories, in: Lecture Notes in Mathematics 86 (Springer-Verlag, 1969) pp. 76–122.
[8] J.L.MacDonald, Coherence of adjoints, associativities, and identities, Arch. Math. 19 (1968) 398–401.
[9] S.MacLane, Natural associativity and commutativity, Rice University Studies 49 (1963) 28–46.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 2 (1971) pp. 141–154.

PAIRS OF RELATIVE COHOMOLOGICAL DIMENSION ONE

C.T.C. WALL

Department of Pure Mathematics, University of Liverpool, England

Received 26 November 1970

We say that a CW pair (W, M) with W connected has cohomological dimension at most n, and write c.d.$(W, M) \leqslant n$, if (i) for the universal cover $p: \widetilde{W} \to W$ and $\widetilde{M} = p^{-1}(M)$, $H_i(\widetilde{W}, \widetilde{M}; \mathbf{Z}) = 0$ for all $i > n$, and (ii) for all coefficient bundles $\mathcal{B}$ over W, $H^{n+1}(W, M; \mathcal{B}) = 0$.

If c.d.$(W, M) \leqslant n$, where $n \geqslant 3$, it is easy to show with the methods of [14] that there exist a complex $L \supset M$ with $\dim(L - M) \leqslant n$ and a homotopy equivalence (which can be chosen simple if W is finite) $L \to W$ (rel M).

However for $n = 1$ this need not be true. For an example, take W contractible and M of type $(\pi, 1)$ where π is one of the groups of [1, 6]. An evident further necessary condition in this case is that for each $x \in M$, $\pi_1(M; x) \to \pi_1(W; x)$ be injective. When this also is satisfied, we will say that (W, M) satisfies D1. We conjecture that given any CW pair (W, M) satisfying D1, there is a homotopy equivalence $W \to L$ (rel M) with $\dim(L - M) \leqslant 1$.

In this paper we first reduce the problem to one in group theory, by showing that it depends only on the fundamental groups involved. We will study this problem, assuming these groups finitely generated. In this case we can reduce the conjecture to another one concerning 'accessibility' of the groups in a sense to be defined in Section 3. A more detailed discussion of this reduction is given at the end of the paper.

Even assuming our conjecture, it is not possible when W is finite to choose the above homotopy equivalence $L \to W$(rel M) to be simple: there is a well-defined obstruction in the Whitehead group of $\pi_1(W)$.

The case when M is empty is equivalent to the result of Stallings [12] and Swan [13] that groups of cohomological dimension 1 are free. We can thus assume $M \neq \emptyset$. A first relativisation of this result is also due to Swan. We will make much use of the results and techniques of these two papers.

§1. Reduction of the problem

In this section, we reduce the problem to a problem in group theory, and show how the general result follows from the case when M is connected.

Write G for $\pi_1(W)$ and $\{H_i: i \in I\}$ for the fundamental groups of the components of M, indexed by $I = \pi_0(M)$. Write $\bigcup_{i\in I} K(H_i, 1)$ for the disjoint union, and embed this in $K(G, 1)$ so as to induce (up to conjugacy) the inclusion maps of the fundamental groups.

Theorem 1. *If* (W, M) *satisfies* D1, *so does* $(K(G, 1), \bigcup_{i\in I} K(H_i, 1))$.

Proof. Let A be a family of arcs in $\widetilde{W}$ disjoint except perhaps at their end points from $\widetilde{M}$ and each other, such that each arc joins two components of $\widetilde{M}$, $\widetilde{M} \cup A$ is connected, and A is minimal with this property. Such a family certainly exists if we assume, as we may, that M contains the 0-skeleton of W; we can take A to be a set of 1-cells. Then $H_i(\widetilde{M} \cup A, \widetilde{M}) = 0 = H_i(\widetilde{W}, \widetilde{M})$ for $i \neq 1$, and by minimality of A,

$$H_1(\widetilde{M} \cup A, \widetilde{M}) \to H_1(\widetilde{W}, \widetilde{M}) = \operatorname{Ker}(H_0(\widetilde{M}) \to \mathbf{Z})$$

is an isomorphism. Hence $H_i(\widetilde{W}, \widetilde{M} \cup A) = 0$ for all i, and since each component of $\widetilde{M}$ is 1-connected so is $\widetilde{M} \cup A$; thus $\widetilde{M} \cup A \subset \widetilde{W}$ is a homotopy equivalence.

Choose an inclusion $M \subset K_M = \bigcup_{i\in I} K(H_i, 1)$ inducing an isomorphism of fundamental groupoids, and with $K_M \cap W = M$. By van Kampen's theorem, $W \subset K_M \cup W$ also induces an isomorphism of fundamental groupoids, so its universal cover contains $\widetilde{W}$: write $(K_M \cup W)^\sim = \widetilde{K}_M \cup \widetilde{W}$. Now a deformation retraction of $\widetilde{W}$ on $\widetilde{M} \cup A$ induces one of $(K_M \cup W)^\sim$ on $\widetilde{K}_M \cup A$. Each component of $\widetilde{K}_M$ is contractible, and A contains just enough arcs to connect them up, so $\widetilde{K}_M \cup A$ is contractible. Hence $K_M \cup W$ is of type $(G, 1)$. The result now follows as the relative homology and cohomology groups of the pairs (W, M) and $(K_M \cup W, K_M)$ coincide.

We now reduce our original problem to a group-theoretic one. If $\dim(W - M) = 1$, then (up to homotopy) W is formed from M by attaching 1-cells. It follows that G is the free product of (conjugates of) the subgroups H_i and a free group. Conversely, assume this. Then we can attach 1-cells to M and map the resulting L to W so as to induce an isomorphism on fundamental groups. By a mapping cylinder construction, suppose $L \subset W$. Then $\widetilde{L}$ is obtained from $\widetilde{M}$ by adding arcs A just as above, so $\widetilde{L} \subset \widetilde{W}$ is a homotopy equivalence. Hence so is the inclusion $L \subset W$.

Say that $(G, \{H_i\})$ satisfies D1 when the H_i are subgroups of G such that $(K(G, 1), \bigcup_{i\in I} K(H_i, 1))$ satisfies D1. We see that our original problem is equivalent to showing that this implies that G is the free product of (conjugates of) the H_i and a free group.

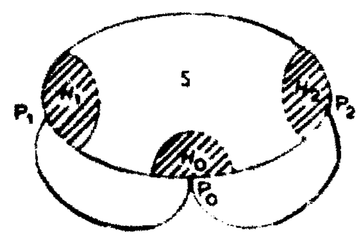
The most natural case of the above is when M is connected, so $\{H_i\}$ is just one subgroup H. The simplification is not inherited by subgroups, however. For let S be

a subgroup of G; regard $K(S, 1)$ correspondingly as a covering space of $K(G, 1)$. The induced covering of $K(H, 1)$ then splits into components $K(S \cap H^g, 1)$, where g runs through representatives of the double cosets $HgS \subset G$; and it follows that $(S, \{S \cap H^g\})$ satisfies D1.

In good cases, the above can be reversed.

Theorem 2. *Let* $(S; H_0, H_1, H_2, \ldots, H_n)$ *satisfy* D1. *Let G be the free product of S and infinite cyclic groups with generators $x_1, \ldots, x_n$. Let H be the subgroup which is the free product of H_0 and the $H_i^{x_i}$. Then (G, H) satisfies* D1.

Proof. Form $\cup K(H_i, 1) \subset K(S, 1)$ as above; choose $P_i \in K(H_i, 1)$, and form a $K(G, 1)$ from $K(S, 1)$ by attaching arcs joining P_0 to P_i $(1 \leqslant i \leqslant n)$. We can then identify $K(H, 1)$ with the union of the $K(H_i, 1)$ and the arcs.



The (relative) cohomology of S is computed from a free chain complex C_* with generators given by lifts to the universal cover of cells of $K(S, 1)$. For (G, H) we can use the same cells and even the same liftings. The boundaries of the cells are unaltered by the reinterpretation, so we have the chain complex $C_* \otimes_{ZS} ZG$. Now for any ZG-module M,

$$\begin{aligned} H^2(K(G, 1), K(H, 1); M) &= H^2(\mathrm{Hom}_{ZG}(C_* \otimes_{ZS} ZG, M)) \\ &= H^2(\mathrm{Hom}_{ZS}(C_*, M)) \\ &= H^2(K(S, 1), \cup K(H_i\, 1); M) = 0\,. \end{aligned}$$

Since condition (i) is immediate by excision, the result follows.

Suppose we can show that for (G, H) satisfying D1, G is the free product of H and a free group: thus we can form a $K(G, 1)$ from $K(H, 1)$ by attaching 1-cells. Suppose G as above; consider the covering space of $K(G, 1)$ corresponding to the subgroup S of G – it is of type $(S, 1)$. The inverse image of $K(H, 1)$ is (as noted above) $\cup K(S \cap H^g, 1)$. The only non-contractible components of this correspond to the nontrivial $S \cap H^g$, and these are just the H_i. Shrinking each contractible one to a point does not alter homotopy type, and we now see that we have a $K(S, 1)$ formed from the $K(H_i, 1)$ for the given subgroups of S by attaching 0-cells and 1-cells. It follows that S is a free product of conjugates of the H_i and a free group.

§2. Cohomology theory

We will now interpret our hypothesis in the more group-theoretic terms which will be needed below. I have been careful to write 'D1' to avoid potential confusion with other interpretations of relative cohomological dimension for groups. For example, it does not mean that the $\mathrm{Ext}^n_{(G,H)}$ groups of Hochschild [8] vanish for $n \geq 2$, though an interpretation may be sought using the 'mixed' Ext^2 groups of Butler and Horrocks [3]. In this paper we will be more down-to-earth.

Write P for the kernel of the map induced by augmentation,

$$\epsilon: \bigoplus_{i\in I}(\mathbf{Z}G \otimes_{\mathbf{Z}H_i} \mathbf{Z}) \to \mathbf{Z}.$$

Lemma 3. *Let $\{H_i : i \in I\}$ be a family of subgroups of G. Then $(G, \{H_i\})$ satisfies* D1 *if and only if P is a projective $\mathbf{Z}G$-module.*

Proof. Write $W = K(G, 1)$, $M = \bigcup_{i\in I} K(H_i, 1)$, with $M \subset W$. Let $p: \widetilde{W} \to W$ be the universal cover, $\widetilde{M} = p^{-1}(\widetilde{W})$. In the exact sequence of $\mathbf{Z}G$-modules

$$H_1(\widetilde{W}) \to H_1(\widetilde{W}, \widetilde{M}) \to H_0(\widetilde{M}) \to H_0(\widetilde{W}) \to H_0(\widetilde{W}, \widetilde{M})$$

the extreme terms vanish, $H_0(\widetilde{W}) \cong \mathbf{Z}$, and we can identify $H_0(\widetilde{M})$ with $\bigoplus_{i\in I}(\mathbf{Z}G \otimes_{\mathbf{Z}H_i} \mathbf{Z})$.

Thus $P \cong H_1(\widetilde{W}, \widetilde{M})$, and by the argument of [14, Lemma 2.1], P is projective.

Conversely, since $H_i(\widetilde{W}, \widetilde{M}) = 0$ for $i \neq 1$, if P is projective the chain complex of $(\widetilde{W}, \widetilde{M})$ is equivalent (in the derived category of complexes of $\mathbf{Z}G$-modules) to one with P as the only nonzero module (see e.g. [15, Theorem 6]), so that $\mathrm{c.d.}(W, M) \leq 1$.

We can rewrite this condition in the case when I is a singleton. Write $I_G = \mathrm{Ker}\,\epsilon$: $\mathbf{Z}G \to \mathbf{Z}$ for the augmentation ideal of G (similarly for H, etc.). Then we have a commutative diagram with exact rows

$$\begin{array}{ccccccccc}
0 & \to & \mathbf{Z}G\cdot I_H & \to & \mathbf{Z}G & \to & \mathbf{Z}G \otimes_{\mathbf{Z}H} \mathbf{Z} & \to & 0\\
 & & \downarrow & & \| & & \downarrow\epsilon & & \\
0 & \to & I_G & \to & \mathbf{Z}G & \to & \mathbf{Z} & \to & 0.
\end{array}$$

If we regard this as a short exact sequence of chain complexes, it has a homology sequence, which boils down to an isomorphism,

$$I_G/\mathbf{Z}G\cdot I_H \to \mathrm{Ker}\,\epsilon = P.$$

Corollary. *If (G, H) satisfies* D1, *I_G splits as the direct sum $\mathbf{Z}G \cdot I_H \oplus P$.*

This is essentially equivalent to a lemma of Swan [13, 4.6]. Note that since $\mathbf{Z}G \otimes_{\mathbf{Z}H} \mathbf{Z}$ is the free Abelian group on cosets of H in G, P can only vanish if $H = G$.

§3. Theory of ends

The classical theory of ends of finitely generated groups is described, for example, in [4, 5]. For our purposes, the following statement will suffice. If G is finite, G has 0 ends. If G is infinite, but $H^1(G; \mathbf{Z}G) = 0$, G has 1 end. If $H^1(G; \mathbf{Z}G) \cong \mathbf{Z}$, G has 2 ends. Otherwise, $H^1(G; \mathbf{Z}G)$ is not finitely generated, and G has ∞ ends. This led to the problem of determining the number of ends of G in terms of its group theoretic structure. This problem was solved by John Stallings. However, the case of 2 ends was characterised earlier by Hopf [9] as that in which G has an infinite cyclic subgroup of finite index. As was shown by myself [16] and Stallings [12], this is equivalent to having a finite normal subgroup with quotient infinite cyclic or dihedral.

To describe the results, we need the concept of free product with amalgamated subgroup, in two forms.

(i) Let A, B, C be groups, $i_0: C \to A$ and $i_1: C \to B$ monomorphisms. Write $A *_C B$ for the quotient of the free product $A * B$ by the normal subgroup generated by elements $i_0(c)\, i_1(c^{-1})$. An expression $G \cong A *_C B$ is called a decomposition of G, and is trivial only if i_0 or i_1 is an isomorphism.

(ii) Let A, C be groups, i_0 and $i_1: C \to A$ monomorphisms.

Write $A *_C$ for the quotient of $A * \mathbf{Z}$ by the normal subgroup generated by elements $t^{-1} i_0(c)\, t i_1(c^{-1})$. An expression $G \cong A *_C$ is a non-trivial decomposition of G.

Theorem 4 (Stallings). *The finitely generated group G has more than one end if and only if G has a non-trivial decomposition $A *_C B$ or $A *_C$, with C finite.*

Sufficiency is easily checked: we remark that the case of 2 ends occurs with a decomposition $A *_C$ when i_0 (hence also i_1) is an isomorphism, and with $A *_C B$ when $|A: i_0(C)| = |B: i_1(C)| = 2$. In these cases there is only one non-trivial decomposition.

For the proof, we refer to [12] (and rely on [2] or [4] to extend the arguments to the finitely generated case). To extend the results to the non-torsion-free case, the crucial step is the following strengthening (due to Stallings) of [12, 4.1]. We give an outline of the argument, since the result is not yet published.

Lemma 5. *Write C for the set of $g \in G$ such that two of $E \cdot gE$ etc. are finite. Then C is a finite subgroup of G.*

Stallings' argument with bipolar structures then leads to a decomposition $G = A *_C B$ or $A *_C$.

Proof. By the argument of [12, 4.1], C is the subgroup of $g \in G$ such that $gP + P$ bounds a finite cochain. For such g, either $gE + E$ or $gE + E^*$ is finite. Let D be the

subgroup of C (of index at most 2) consisting of g with $gE + E$ finite. By [12, 4.6], the set of g such that gP meets P is finite. But if $g \in D$ and gP does not meet P, then [12, 2.3] either $gE \subset E$ or $E \subset gE$. But now by induction $g^r E \subset E$ (or $E \subset g^r E$) for all natural numbers r, so g has infinite order. Assuming G has infinitely many ends, this contradicts [12, 3.5].

In the torsion-free case, since $A *_{\{1\}} \cong A * \mathbf{Z}$, the conclusion of Theorem 5 can be paraphrased by saying that G is either a free product or infinite cyclic. Since by Grushko's theorem [11, 7] G cannot be infinitely decomposed as a free product it follows by induction that any finitely generated torsion-free group is the free product of a free group and groups with one end.

This argument is still lacking in the general case, and constitutes the one gap in the development of this paper. We now formalise the problem. Say that G is 0-accessible if it has at most 2 ends, and inductively that it is n-accessible if it has the form $A *_C B$ or $A *_C$ with C finite and A, B $(n-1)$-accessible (trivial decompositions not excluded).

Conjecture. Every finitely generated group is n-accessible for some n.

The validity of our solution to the problem raised at the beginning of the paper depends on the accessibility of certain groups related to G. Certainly all is well if G is torsion free, but in that case a much simpler argument suffices anyway.

§4. Subgroups of amalgamated free products

Since Theorem 4 is our main technical tool, we shall need information about free products with amalgamation, particularly about their subgroups. For ordinary free products one has the Kuroš subgroup Theorem, which is well-known to follow easily from covering space arguments. We will use these methods here also, having a taste for geometry. The arguments can also be formulated algebraically, using groupoids, as in Higgins [7], adapting the geometrical argument of [11].

Before starting the general discussion, we make the following observation, which will be needed later.

Remark. If $G = A *_C B$ or $A *_C$, with C finite, is not a free product, then neither are A and B.

If, in the first case, $A = D * E$ then by Kuroš Theorem C is conjugate to a subgroup of D or E, say E; conjugating, we may suppose $C \subset E$. But then

$$G = (D * E) *_C B = D * (E *_C B)$$

is a free product, a contradiction. In the second case we either have the same contra-

diction or can suppose $i_0(C) \subset D$, $i_1(C) \subset E$. But then G is the free product of $t^{-1}Dt *_{i_1(C)} E$ and the infinite cyclic group generated by t.

Our method consists in the following. Suppose $G = A *_C B$ or $A *_C$. Then we can form an Eilenberg-MacLane space $K(G, 1)$ from $K(A, 1) \cup K(B, 1)$ or $K(A, 1)$ by attaching $K(C, 1) \times I$ along $K(C, 1) \times \{0, 1\}$ which is injected via i_0 and i_1. Now for any subgroup S of G we consider the covering space of $K(G, 1)$ corresponding to S. It is decomposed into parts corresponding to A, B and C. Indeed, let H be any subgroup of G with $K(H, 1) \subset K(G, 1)$. Then the induced covering of $K(H, 1)$ breaks up into components which correspond bijectively to double cosets HgS: the corresponding component has fundamental group $H^g \cap S$.

Moreover, this description is natural: suppose we have $K(C, 1) \subset K(A, 1) \subset K(G, 1)$. Then the components of the S-covering space corresponding to double cosets $Cg_1 S$, $Ag_2 S$ will be incident if and only if $Cg_1 S \subset Ag_2 S$.

Thus starting from $G = A *_C B$ or $A *_C$ and $S \subset G$, we have a $K(S, 1)$ which looks like a graph, with (at each vertex) $K(A^g \cap S, 1)$ or $K(B^g \cap S, 1)$ and edges of the form $K(C^g \cap S, 1) \times I$: the incidences being determined as above. Unfortunately, the graph so obtained is usually infinite and we must work to reduce it to a finite subgraph. To illustrate the technique, we now use it to prove one simple result.

Theorem 6. *Let G be a finitely generated subgroup of a group $A *_C B$ or $A *_C$. Then either G is conjugate to a subgroup of A or of B, or G has a non-trivial decomposition $X *_Z Y$ or $X *_Z$, with Z conjugate to a subgroup of C.*

Proof. Represent i_0 by an inclusion $K(C, 1) \times 0 \subset K(A, 1)$; similarly for i_1. Then $K(A, 1) \cup K(C, 1) \times I \cup K(B, 1)$, resp. $K(A, 1) \cup K(C, 1) \times I$ is a space of type $(A *_C B, 1)$, resp. $(A *_C, 1)$. Look at the covering space corresponding to the subgroup G. The preimage of $K(C, 1) \times I$ breaks up into components $W_\alpha \times I$, say.

If any of these fail to disconnect $K(G, 1)$, write X for the fundamental group of its complement, $Z = \pi_1(W_\alpha)$: then $G = X *_Z$ as required.

If $W_\alpha \times I$ separates $K(G, 1)$ into two components, with fundamental groups X and Y, and $Z = \pi_1(W_\alpha)$, then $G = X *_Z Y$ is a decomposition as required, unless it is trivial – i.e. Z maps onto one of X, Y.

We may thus suppose that for each W_α we obtain a trivial decomposition of G. Now since G is finitely generated, we can choose a finite set of loops which generate it; these will lie in a finite (connected) union L of components U_β of the preimage of $K(A, 1)$; V_γ of the preimage of $K(B, 1)$; and $W_\alpha \times I$. Since i_0, i_1 are injective, so are the maps of fundamental groups induced by inclusions $W_\alpha \times 0 \to U_\beta$, $W_\alpha \times 1 \to V_\gamma$ (or U_β). It follows that $\pi_1(L) \to G$ is injective, and hence an isomorphism.

We will now make an induction on the number n_L of α such that $W_\alpha \times I \subset L$. If $n_L = 0$, $L = U_\beta$ or V_γ, and $G = \pi_1(L)$ is conjugate to a subgroup of A or B, as asserted. Otherwise, choose $W_\alpha \times I \subset L$; let it separate $L = L_0 \cup (W_\alpha \times I) \cup L_1$. Since the resulting decomposition is trivial we can choose one of L_0 and L_1, say L_0,

so that $\pi_1(L_0) = \pi_1(L) = G$. Now we can apply the induction hypothesis to L_0. The result follows.

§5. The case when G has one end

We are now ready to start work on our main problem: in this section, after some preliminaries, we prove the result in the cases when G has at most 2 ends.

Lemma 7. *Let $(G, \{H_i\})$ satisfy* D1; *let $g^{-1}hg = k$ with $g \in G$, $1 \neq h \in H_i$, $k \in H_j$. Then $i = j$ and $g \in H_i$.*

Proof. Represent h, k by maps

$$S^1 \times 0 \to K(H_i, 1), \quad S^1 \times 1 \to K(H_j, 1)\,.$$

The hypothesis implies that we can extend to a map

$$f: S^1 \times I \to K(G, 1) = W, \quad \text{say},$$

where the path of $1 \times I$ represents g. Let $\bar{W}$ be the covering space corresponding to the (non-trivial) cyclic subgroup S generated by h, and $\bar{M}$ the corresponding covering of $\cup K(H_i, 1)$. Our homotopy f lifts to a map $F: S^1 \times I \to \bar{W}$, since $f \mid (S^1 \times 0)$ does by construction. Since* $H_2(\bar{W}, \bar{M}) = 0$, F represents the zero homology class, so $F(S^1 \times 0 \cup S^1 \times 1)$ is nullhomologous in $\bar{M}$. But $F(S^1 \times 0)$ is not, so $F(S^1 \times 1)$ must lie in the same component of $\bar{M}$. It follows that $i = j$, and that g, which lifts to a path joining this component to itself – the component corresponding to the double coset $H_i 1 S = H_i$ – belongs to H_i.

Corollary 1. *If $i \neq j$, $H_i \cap H_j = \{1\}$. The nontrivial H_i are distinct subgroups of G.*

Corollary 2. *Suppose H_1 contains a nontrivial normal subgroup of G. Then $H_1 = G$ and $H_i = \{1\}$ for $i \neq 1$.*

These follow immediately from the lemma.

Proposition 8. *Let F be a finite subgroup of G. Then for some $g \in G$, $i \in I$, we have $g^{-1}Fg \subset H_i$.*

* It follows from [15] that if (W, M) satisfies D1, the relative homology and cohomology groups, with respect to any coefficient bundle, vanish in dimensions $\geqslant 2$.

Proof. First suppose F of prime order and the result false. Then no conjugate of F meets any H_i. Thus the covering space $\tilde{W}$ of $K(G, 1)$ corresponding to F induces a covering $\tilde{M}$ of $\bigcup K(H_i, 1)$ each component of which is contractible. But then

$$H^2(\tilde{W}, \tilde{M}; \boldsymbol{Z}) = H^2(F; \boldsymbol{Z}) \neq 0 \,,$$

a contradiction.

Now use induction on $|F|$. Let M be a maximal subgroup: since $|F|$ is not prime, $1 < |M| < |F|$. By induction, we can write $g^{-1}Mg \subset H_i$. If now for some $x \in (F - M)$, M meets $x^{-1}Mx$ nontrivially then by Lemma 7 $g^{-1}xg \in H_i$: now, since M is maximal, $g^{-1}Fg \subset H_i$. But if $x^{-1}Mx \cap M = \{1\}$ for all $x \in (F - M)$, F is a Frobenius group and so has a proper normal subgroup. If we use this instead of M, we obtain the desired result.

Corollary 1. *If G is finite and $(G, \{H_i\})$ satisfies* D1, *then some $H_i = G$ and the rest are trivial.*

Corollary 2. *If G has two ends and $(G, \{H_i\})$ satisfies* D1, *either some $H_i = G$ and the rest are trivial, or $G \cong \boldsymbol{Z}$, all H_i are trivial, or $G \cong \boldsymbol{Z}_2 * \boldsymbol{Z}_2$, two H_i are conjugate to the free factors, and the rest are trivial.*

For G has a finite normal subgroup F with G/F isomorphic to $\boldsymbol{Z}$ or to $\boldsymbol{Z}_2 * \boldsymbol{Z}_2$. If F is nontrivial, by Proposition 8 we have $F \subset H_i$, and the result now follows by Corollary 2 to Lemma 7. Similarly if $G = \boldsymbol{Z}$ and H_i is nontrivial or if $G \cong \boldsymbol{Z}_2 * \boldsymbol{Z}_2$ and H_i has order greater than 2 (when it contains a nontrivial normal subgroup), we find $H_i = G$. Otherwise, by Proposition 8, the two factors $\boldsymbol{Z}_2$ are conjugate to distinct H_i, H_j; as there are no more conjugacy classes of subgroups of order 2, any further H_k are trivial.

Theorem 9. *Let G have one end and $(G, \{H_i\})$ satisfy* D1. *Then some $H_i = G$, and further H_j are trivial.*

Proof. We have epimorphisms (where $n = |I|$)

$$n\boldsymbol{Z}G \xrightarrow{q} \bigoplus_{i \in I} (\boldsymbol{Z}G \underset{\boldsymbol{Z}H_i}{\otimes} \boldsymbol{Z}) \xrightarrow{\epsilon} \boldsymbol{Z} \,;$$

write $(P, p) = \operatorname{Ker} \epsilon$ and $(J, j) = \operatorname{Ker} q$. Since by hypothesis P is projective, there exists $s: P \to n\boldsymbol{Z}G$ with $qs = p$. Choose a $\boldsymbol{Z}G$-free resolution

$$R \xrightarrow{u} F \xrightarrow{v} J \to 0 \,.$$

Then the following is also exact:

$$R \xrightarrow{(u,0)} F \oplus P \xrightarrow{(jv,s)} n\boldsymbol{Z}G \xrightarrow{\epsilon q} \boldsymbol{Z} \to 0 \,.$$

Since G has one end, $H^0(G; \mathbf{Z}G)$ and $H^1(G; \mathbf{Z}G)$ vanish, so the dual sequence is exact too:

$$R^* \xleftarrow{(u^*,0)} F^* \oplus P^* \xleftarrow{(v^*j^*,s^*)} n\mathbf{Z}G \leftarrow 0\,.$$

The middle map gives an isomorphism of $\operatorname{Ker} u^* \oplus P^*$ with $n\mathbf{Z}G$: denote its restriction by

$$k^*: P^* \to n\mathbf{Z}G\,.$$

Then $s^*k^* = 1$ and $v^*j^*k^* = 0$.

Since G is a finitely generated group, the kernel of the augmentation $\mathbf{Z}G \to \mathbf{Z}$ is a finitely generated $\mathbf{Z}G$-module. Hence so are the kernel of ϵq, and its quotient P. Thus P is reflexive. Hence dualising k^* gives a map k: $n\mathbf{Z}G \to P$, and $ks = 1$. Also $kjv = 0$, as it factors through its double dual

$$F \to F^{**} \xrightarrow{k^{**}j^{**}v^{**} = (v^*j^*k^*)^* = 0} P^{**} = P\,.$$

As v is surjective, $kj = 0$. Hence k factors through q, $k = lq$ say. Now $1_P = ks = lqs = lp$, so l splits p. Hence also ϵ is split, say by m.

Since $m \neq 0$, for some i there is a nontrivial $\mathbf{Z}G$-homomorphism m': $\mathbf{Z} \to \mathbf{Z}G \otimes_{\mathbf{Z}H_i} \mathbf{Z}$. But each element of G/H_i must appear with the same nonzero coefficient in $m'(1)$. Thus H_i has finite index in G, and so contains a normal subgroup of finite index, necessarily nontrivial. The result now follows from Lemma 7, Corollary 2.

§6. The induction step

Let G be a finitely generated group. We will say that G satisfies (P) if it is neither a free product nor infinite cyclic: equivalently, if it has no non-trivial decomposition $A *_{\{1\}} B$ or $A *_{\{1\}}$. We say that G satisfies (Q) if whenever $(G, \{H_i\})$ satisfies D1, we have all H_i trivial except one which equals G. A special case of our main conjective is that (P) implies (Q). In the previous section we have proved this for groups G with at most 2 ends.

Theorem 10. *Let G satisfy* (P) *and have the form $A *_C B$ or $A *_C$, with C finite. If A and B satisfy* (Q), *so does G.*

Note that by a remark in Section 4, A and B satisfy (P) if G does. It thus follows by induction that

Corollary. *If G is accessible and satisfies* (P), *it also satisfies* (Q).

Thus we have proved that (P) implies (Q), modulo the conjective that all finitely generated groups are accessible.

Proof of Theorem 10. Suppose $(G, \{H_i\})$ satisfies D1. Write $H_i = *H_{ij} * F$, where F is free and the H_{ij} are not free products nor infinite cyclic (this is possible by Gruško's Theorem). Then $(H_i, \{H_{ij}\})$ satisfies D1, hence so does $(G, \{H_{ij}\})$. If we can show that some $H_{ij} = G$, it follows that $H_i = G$, hence the result. So it suffices to proceed on the assumption that the H_i satisfy (P).

By Proposition 8, C is conjugate to a subgroup of some H_i, so we may suppose $C \subset H_1$. By Lemma 7, if $i > 1$, H_i meets no conjugate of C, and H_1 meets C^g only if $g \in H_1$. Hence by Theorem 6, since H_i satisfies (P), it is conjugate to a subgroup of A or of B, if $i > 1$. For H_1 itself, we refer to the proof of Theorem 6, in the notation there used. We have $\pi_1(W_\alpha)$ nontrivial for only a single α: since H_1 satisfies (P) we can then reduce n_L to 1. Hence $H_1 = H_1^A *_C H_1^B$, say, with $H_1^A \subset A$ and $H_1^B \subset B$: or, if $G = A *_C$, we may have $H_1 = H_1^A *_C$ with $H_1^A \subset A$.

It thus follows that we can choose an embedding of $M = \bigcup K(H_i, 1)$ in

$$W = K(G, 1) = K(A, 1) \cup K(C, 1) \times I \cup K(B, 1)$$

$$= W_A \cup W_C \times I \cup W_B \text{ , say,}$$

such that M has the form $M_A \cup W_C \times I \cup M_B$ with $M_A \subset W_A$, $M_B \subset W_B$ (of course, if $G = A *_C$, W_B is to be omitted). But the pairs (W_A, M_A), (W_B, M_B) also satisfy D1. For M_A is a union of certain $K(H_i, 1)$ with the H_i conjugated to lie in A, and of $K(H_1^A, 1)$, or possibly (when $G = A *_C$) also a $K(H_1^{A'}, 1)$. Hence the injectivity condition is satisfied. As to the other, the condition on homology follows by excision, and if R is any $\mathbf{Z}A$-module, let R^G be the induced $\mathbf{Z}G$-module. Then R is a $\mathbf{Z}A$-direct summand of R^G, and

$$H^2(W, M; R^G) \cong H^2(W_A, M_A; R^G) \oplus H^2(W_B, M_B, R^G)$$

by excision (Swan needs a lengthy algebraic development for the analogous result [13, 2.3], but it follows at once from the topological interpretation); since the left-hand side vanishes, so do $H^2(W_A, M_A; R^G)$ and its direct summand $H^2(W_A, M_A; R)$.

But A satisfies (Q). Since $H_1^A \supset C$ is already nonzero, it follows that $H_1^A = A$ and there are no further $H_i \subset A$: thus in particular when $G = A *_C$ we cannot have $H_1 = H_1^A *_C H_1^{A'}$. Thus if $G = A *_C$, $H_1 = H_1^A *_C = A *_C = G$. If $G = A *_C B$ then, similarly, $H_1^B = B$ and again $H_1 = G$. This proves the theorem.

We now use a similar argument for groups which do not satisfy (P).

Theorem 11. *Let $(G, \{H_i\})$ satisfy* D1. *Decompose G and the H_i as free products of indecomposable groups; ignore free factors isomorphic to $\{1\}$ or $\mathbf{Z}$. Then the set of free factors of the H_i coincides (up to conjugacy in G) with the set of free factors of G, provided these satisfy* (Q).

Proof. As for Theorem 10, we may assume that the H_i satisfy (P). Then if $G = A * B$, we may assume by Theorem 6 that each H_i is contained in A or in B. It follows, as above, that $(A, \{H_i\})$ satisfies D1. The result thus follows by induction on the number of indecomposable free factors of G (note that in the case $G = \mathbf{Z}$ nothing is asserted).

Corollary. *Let (G, H) satisfy* D1 *and H satisfy* (P). *Suppose G accessible. Then G is the free product of H and a free group.*

For decompose G as in the Theorem. By an argument in the next section, each factor is accessible, and hence by Theorem 10, Corollary, satisfies (Q). The result thus follows from the Theorem.

§7. Conclusion

We will deduce our main result from those of the preceding section by a trick due to Swan [13, 5.1]. Let (G, H) satisfy D1. Write

$$L = G *_H (H \times \mathbf{Z}) = G *_H .$$

If H is finite, we will see that accessibility of L implies that of G.

Theorem 12. *Let (G, H) satisfy* D1 *and $G *_H$ be accessible. Then G is the free product of H and a free group.*

Proof. If H is finite, it satisfies (P) and the result follows from Theorem 11, Corollary. Otherwise, we form L as above. As in Theorem 2, it follows by excision that $(L, H \times \mathbf{Z})$ satisfies D1. Since H is infinite, $H \times \mathbf{Z}$ has only one end, and hence satisfies (P). Again by the corollary to Theorem 11, we have a free product decomposition $L = (H \times \mathbf{Z}) * F$, with F a free group. Factoring out the normal subgroup generated by $\mathbf{Z}$, it follows that $G = H * F$, as stated.

It remains to comment a little on the problem of accessibility. First, the argument works without a hitch if G is torsion-free; and for this it suffices by Proposition 8 that the H_i be so. However, in this case the result follows immediately by combining (4.2) and (5.1) of [13].

Secondly, it is possible to improve the hypothesis of Theorem 12 somewhat. I

have no results assuming merely that H (or the H_i) are accessible, but an earlier (and much more complicated) proof of Theorem 12 using only accessibility of G. The idea was to use a decomposition of L (which has ∞ ends by results of Swan) and study exactly how it decomposed G; then to use induction on the rank of $H_1(W, M; \mathbf{Z})$. The induction base came from combining the Corollary to Theorem 11, which implies that W can be obtained (up to homotopy) from M by adding 1-cells and 2-cells, with a result of Kaplansky [10]. We can then combine Theorems 2 and 12 to obtain

Conclusion. *Suppose* $(G; \{H_i\})$ *satisfies* D1, *all groups finitely generated, G accessible. Then G is the free product of the* H_i *and a free group.*

Finally, it remains to check two earlier assertions about accessibility. Now the definition shows inductively that G is accessible iff we can build a $K(G, 1)$ modelled on a finite graph, with each vertex replaced by a $K(H, 1)$ (H with 0 or 1 end) and each edge by a $K(F, 1) \times I$, F finite. If G is a free product $X * Y$, each F, H is (essentially uniquely) conjugable into a subgroup of X or Y, by Theorem 6. If we cut the edges corresponding to trivial F, the components of the remainder thus correspond to subgroups of X or Y. Hence by rearranging the cut edges, we can form a $K(X, 1)$ as a subgraph of the graph for $K(G, 1)$ so X is accessible. The argument that $G *_H$ accessible, H finite, implies G accessible is similar.

References

[1] G. Baumslag and K.W. Gruenberg, Some reflections on cohomological dimension and freeness, J. Alg. 6 (1967) 394–409.

[2] G.M. Bergman, On groups acting on locally finite graphs, Ann. of Math. 88 (1968) 335–340.

[3] M.C.R. Butler and G. Horrocks, Classes of extensions and resolutions, Phil. Trans. Roy. Soc. A254 (1961) 155–222.

[4] M.J. Dunwoody, The ends of finitely generated groups, J. Alg. 12 (1969) 339–344.

[5] D.B.A. Epstein, Ends, pp. 110–117, in: Topology of 3-manifolds and related topics, ed. M.K. Fort Jr. (Prentice-Hall, 1962).

[6] D.B.A. Epstein, A group with zero homology, Proc. Camb. Phil. Soc. 64 (1968) 599–601.

[7] P.J. Higgins, Grushko's theorem, J. Alg. 4 (1966) 365–372.

[8] G. Hochschild, Relative homological algebra, Trans. Am. Math. Soc. 82 (1956) 246–269.

[9] H. Hopf, Enden offene Räume und unendlich diskontinuierliche Gruppen, Comm. Math. Helv. 16 (1943) 81–100.

[10] M. Susan Montgomery, Left and right inverses in group algebras, Bull. Am. Math. Soc. 75 (1969) 539–540.

[11] J.R. Stallings, A topological proof of Gruško's Theorem on free products, Math. Z. 90 (1965) 1–8.

[12] J.R. Stallings, On torsion free groups with infinitely many ends, Ann. of Math. 88 (1968) 312–334.

[13] R.G. Swan, Groups of cohomological dimension one, J. Alg. 12 (1969) 585–610.

[14] C.T.C. Wall, Finiteness conditions for C.W. complexes, Ann. of Math. 81 (1965) 56–69.
[15] C.T.C. Wall, Finiteness conditions for C.W. complexes II, Proc. Roy. Soc. A 295 (1966) 129–139.
[16] C.T.C. Wall, Poincaré complexes I, Ann. of Math. 86 (1967) 213–245.

JOURNAL OF PURE AND APPLIED ALGEBRA Volume 1, No. 2 (1971) pp. 155–158.

ON A TOPOLOGY FOR A GALOIS SYSTEM OF FIELDS AND AUTOMORPHISMS

T. SOUNDARARAJAN *

Department of Mathematics, Madurai University, Madurai-2, Tamilnadu, S. India

Received 12 February 1971

Let E be a field and G be a group of automorphisms of E. For any subset X of E, let G_X denote the set of all $\sigma \in G$ such that $\sigma x = x$ for every $x \in X$; G_X is a subgroup of G. For any subset S of G, let $I(S)$ denote the set of all $x \in E$ such that $\sigma x = x$ for every $\sigma \in S$; $I(S)$ is a subfield of E. It is obvious that $X \subset I(G_X)$ and $S \subset G_{I(S)}$.

For any subfield F of E, the following two conditions are evidently equivalent:

(i) $F = I(H)$ for some subgroup H of G;

(ii) $F = I(G_F)$.

The subfields F that satisfy these conditions are said to be *Galois closed relative to G*. The set of all of them will be denoted by $\mathcal{F}(G)$. The field $K = I(G)$ is an element of $\mathcal{F}(G)$, and every element of $\mathcal{F}(G)$ is an extension of K.

For any subgroup H of G, the following two conditions are evidently equivalent:

(i) $H = G_F$ for some subfield F of E;

(ii) $H = G_{I(H)}$.

The subgroups H that satisfy these conditions are said to be Galois closed in G. The set of all of them will be denoted by $\mathcal{G}(G)$.

Proposition. *The formula $F \mapsto G_F$ defines a mapping $\mathcal{F}(G) \to \mathcal{G}(G)$, the formula $H \mapsto I(H)$ defines a mapping $\mathcal{G}(G) \to \mathcal{F}(G)$, and these two mappings are inverse to each other.*

This is obvious. What is not obvious is how to characterize the subfields and subgroups that are Galois closed. When E is a differential field and G is a group of differential automorphisms this is a problem raised by Kaplansky [1, p. 18]; when, in

* The author is very thankful to Professor E.R. Kolchin for a full revision of an earlier draft. The author offers his best thanks to Professor M. Venkataraman for his encouragement and guidance in the preparation of this paper.

addition, $\mathcal{F}(G)$ consists of all differential fields between K and E, this is a problem mentioned earlier by Kolchin [2, p. 29].

The purpose of this note is to provide a topological characterization of $\mathcal{G}(G)$ and $\mathcal{F}(G)$.

The set of all subsets of G of the form

$$G - sG_{K(x)} \qquad (s \in G, \quad x \in E)$$

is a subbase for a topology T_G on G. For any $t \in G$, left resp. right translation by t maps $G - sG_{K(x)}$ onto $G - tsG_{K(x)}$ resp. $G - sG_{K(x)}t = G - st \cdot t^{-1}G_{K(x)}t = G - stG_{K(t^{-1}x)}$, it follows that both these translations are homeomorphisms of G onto G. Similarly, the symmetry mapping $t \mapsto t^{-1}$ of G maps $G - sG_{K(x)}$ on the set $G - G_{K(x)}s^{-1} = G - s^{-1}G_{K(sx)}$, and hence is a homeomorphism, too.

For any $x, y \in E$ let $l_{x,y}$ denote the mapping $E \to E$ defined by the formula $l_{x,y}(z) = xz + y$. The set of all subsets of E of the form $E - l_{x,y}(I(s))$ $(x, y \in E$, $s \in G)$ is a subbase for a topology T'_G on E. When $x \neq 0$ then $l_{x,y}$ is a homeomorphism of E onto E.

Theorem. (a) *A subgroup H of G is Galois closed in G if and only if H is a closed set for the topology T_G.*

(b) *A subfield F of E is Galois closed relative to G if and only if $F \supset K$ and F is a closed set for the topology T'_G.*

The proof depends on two lemmas.

Lemma 1. *Let G be any group, $G_1, \ldots, G_m$ be finitely many distinct subgroups of G, $n_1, \ldots, n_m$ be nonzero natural numbers, and s_{ij} be elements of G $(1 \leq i \leq m$, $i \leq j \leq n_i)$. If the $n_1 + \ldots + n_m$ cosets $s_{ij}G_i$ cover G but no fewer of them do then $G_1 \cap \ldots \cap G_m$ is of finite index in G.*

This was proved by B.H. Neumann [3]. Also a proof can be found in [4].

Lemma 2. *Let E be any field F, $F_1, \ldots, F_n$ be finitely many subfields of E, and $x_1, y_1, \ldots, x_n, y_n$ be elements for E. If $F \subset \bigcup_{1 \leq j \leq n} l_{x_j,y_j}(F_j)$ then either F is finite or $F \subset F_j = l_{x_j,y_j}(F_j)$ for some j.*

Proof. Let F be infinite. Discarding superfluous terms $l_{x_j,y_j}(F_j)$, we may suppose that F is not contained in the union of any $n-1$ of the sets $l_{x_j,y_j}(F_j)$. For each j the coset $l_{x_j,y_j}(F_j) = x_jF_j + y_j$ of the subgroup x_jF_j of the additive group E contains an element of F, so that we may suppose that $y_j \in F$. Then

$$F = \bigcup_{1 \leq j \leq n} ((F \cap x_jF_j) + y_j),$$

so that by Lemma 1 $F \cap x_1F_1$ has some finite index r in F (whence in particular $x_1 \neq 0$). We claim $x_1 \in F_1$. Indeed, $F \cap x_1F_1$ is infinite and hence contains $r+1$ distinct nonzero elements $x_1a_0, x_1a_1, \ldots, x_1a_r$; two of the $r+1$ products $x_1a_0 \cdot x_1a_k$ $(0 \leqslant k \leqslant r)$ must lie in the same coset of $F \cap x_1F_1$ in F, so that for two distinct indices k, k' we have $x_1^2a_0(a_k - a_{k'}) \in F \cap x_1F_1 \subset x_1F_1$, whence $x_1 \in F_1$, as claimed. $F \cap F_1$ is of index r in F. For any $x \in F$, $x \neq 0$, $l_{x,0}$ maps the group F isomorphically onto itself and maps $F \cap F_1$ onto $F \cap xF_1$, so that $F \cap xF_1$ has index r in F; by the claim just established, then $x \in F_1$. Therefore $F \subset F_1$ and (because $x_1, y_1 \in F_1$) $F_1 = l_{x_1,y_1}(F_1)$.

We now prove the theorem:

(a) If H is Galois closed then, $H = G_{I(H)} = \bigcap_{x \in I(H)} G_{K(x)}$, so that H is closed for the topology T_G. Conversely, let H be closed for T_G and let $s \in G_{I(H)}$. We must show that $s \in H$, and to do this it suffices to show that every open set $G - \bigcup_{1 \leqslant j \leqslant n} s_jG_{K(x_j)}$ containing s intersects H, that is, to show that

$$s \notin \bigcup s_jG_{K(x_j)} \Rightarrow H \not\subset \bigcup s_jG_{K(x_j)}\,.$$

We assume that $H \subset \bigcup s_jG_{K(x_j)}$ and force a contradiction. Discarding any of the cosets $s_jG_{K(x_j)}$ that are superfluous, we may suppose that no coset $s_jG_{K(x_j)}$ can be omitted. Then each of these cosets contains an element of H, so we can take $s_j \in H$ and write $H = \bigcup s_j(H \cap G_{K(x_j)})$. Lemma 1 then shows that H is contained in the union of finitely many left cosets of the intersection $\bigcap G_{K(x_j)} = G_{K(x_1,\ldots,x_n)}$. Therefore each x_j has only finitely many distinct images tx_j $(t \in H)$ and hence is separably algebraic over $I(H)$. The extension F of $I(H)$ generated by the elements tx_j $(t \in H,\ 1 \leqslant j \leqslant n)$ is normal separably algebraic and of finite degree. The elements of H restrict to automorphisms of F over $I(H)$; since the fixed field of the group of these restrictions is $I(H)$, the group of restrictions is the Galois group of F over $I(H)$. As the restriction of s to F is also an element of this Galois group, some $t \in H$ coincides with s on F, so that $tx_j = sx_j$, and $s_j^{-1}tx_j = s_j^{-1}sx_j$ $(1 \leqslant j \leqslant n)$. Since $s \notin s_jG_{K(x_j)}$ it follows that $t \notin s_jG_{K(x_j)}$ $(1 \leqslant j \leqslant n)$. This contradicts the assumption that $H \subset \bigcup s_jG_{K(x_j)}$, and completes the proof of part (a).

(b) If F is Galois closed then $F \supset K$ and $F = I(G_F) = \bigcap_{s \in G_F} I(s)$, so that F is closed for the topology T'_G. Conversely let F be closed for T'_G with $F \supset K$. When F is finite then K is finite too, and G_F is a normal subgroup of G of finite index: the restrictions to $I(G_F)$ of the elements of G form a finite group G' of automorphisms of $I(G_F)$ with fixed field K, so that $I(G_F)$ is a finite extension of K and G' is its Galois group; since $K \subset F \subset I(G_F)$ and since evidently the identity is the only element of G' that leaves invariant every element of F, it follows that $F = I(G_F)$ and F is Galois closed relative to G. Therefore we may suppose that F is infinite. Let $x \in I(G_F)$. We must show that $x \in F$, and to do this it suffices to show that every open set $E - \bigcup_{1 \leqslant j \leqslant n} l_{x_j,y_j}(I(s_j))$ that contains x intersects F, that is, to show that

$$x \notin \bigcup l_{x_j, y_j}(I(s_j)) \Rightarrow F \not\subset \bigcup l_{x_j, y_j}(I(s_j)) .$$

We assume that

$$F \subset \bigcup l_{x_j, y_j}(I(s_j))$$

and force a contradiction. Then (by Lemma 2) $F \subset I(s_k) = l_{x_k, y_k}(I(s_k))$ for some index k, so that $s_k \in G_F$. Since $x \in I(G_F)$, this implies that $x \in I(s_k) = l_{x_k, y_k}(I(s_k))$. This contradicts the condition $x \notin \bigcup_{x_j, y_j}(I(s_j))$, and completes the proof of the theorem.

Remarks. 1. In the special case in which E is a normal separably algebraic extension of K and G is the group of all automorphisms of E over K, it is easy to see that T_G coincides with the Krull topology on G; despite the fact that $\mathcal{F}(G)$ is then the set of all intermediate fields, T'_G is not the discrete topology on E (even when G is finite).

2. In general, G is not a topological group relative to T_G, because multiplication on G is not continuous as a function of two variables when G^2 is given the product topology.

3. In the special case in which K is a differential field of characteristic 0 with algebraically closed field of constants C and E is a strongly normal extension of K and G is the group of all differential automorphisms of E over K, then G has a canonical structure of algebraic group defined over C, $\mathcal{F}(G)$ is the set of all intermediate differential fields, and a subgroup of G is closed relative to T_G if and only if it is closed relative to the Zariski C-topology on G. Nevertheless, T_G is distinct from the Zariski C-topology (the latter being finer).

References

[1] I. Kaplansky, Introduction to differential algebra (Hermann, Paris, 1957).

[2] E.R. Kolchin, Algebraic matric groups and the Picard-Vessiot theory, Ann. of Math. 49 (1948) 1–42.

[3] B.H. Neumann, Groups covered by finitely many cosets, Publ. Math. Debrecen 3 (1954) 227–242.

[4] T. Soundararajan and K. Venkatachaliengar, On the Krull-Galois theory for nonalgebraic extension fields, Bull. Australian Math. Soc. 4 (1971) 367–387.

[5] T. Soundararajan, Galois theory for general extension fields, J. Reine Angew. Math. 241 (1970) 49–63.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 2 (1971) pp. 159–178.

BETWEENNESS, ORDERS AND INTERVAL GRAPHS

Peter C. FISHBURN *)

The Institute for Advanced Study, Princeton, N.J. 08540, USA

Received 11 December 1970

§ 1. Introduction

Let BT be a ternary relation on a set X, and interpret $(x, y, z) \in \mathrm{BT}$ to mean that y is strictly ** between x and z. We shall write

$$xyz \quad \text{iff} \quad (x, y, z) \in \mathrm{BT}$$

$$\text{not} \quad xyz \quad \text{iff} \quad (x, y, z) \notin \mathrm{BT}\,.$$

A binary relation $<$ on X *agrees with* BT if and only if

$$xyz \quad \text{iff} \quad (x < y < z \quad \text{or} \quad z < y < x)$$

for all $x, y, z \in X$, where $x < y < z$ means that $x < y \;\&\; y < z \;\&\; x < z$. If $<$ is not transitive then $x < y < z$ need not be equivalent to $x < y \;\&\; y < z$.

This paper examines a hierarchy of four types of strict partial orders (asymmetric, transitive), and for each gives conditions on BT that are necessary and sufficient for an agreeing order of the specific type. The least general order considered is a linear order. As noted later in this section, there are well over a dozen extant axiom sets for BT that characterize linear order. The most general order of the four in our hierarchy is an interval order, which has the property that when X is countable it is possible to map each element into a closed real interval such that, for all $x, y \in X$, the interval for x lies wholly to the left (on the line) of the interval for y iff $x < y$. Interval orders are closely related to interval graphs, as we shall observe in the next section. Between interval orders and linear orders we encounter semiorders and weak orders, which will be discussed further in Section 2. Section 3 presents the agreeing theorems.

* The author is grateful to Fred Roberts for sharing his work on betweenness, and to Duncan Luce for introduction to the sizable literature of betweenness. The work was supported by a grant from the Alfred P. Sloan Foundation to the Institute for Advanced Study.

** The strictness interpretation will be relaxed in parts of this section, but all later sections adhere to it unless explicitly noted otherwise.

Ternary betweenness relations, both strict (where not aab) and nonstrict (where aab), were probably used first as primitives in axiomatizations of geometry. According to Freudenthal [1], Pasch [2, 3] was the first to rigorously axiomatize betweenness with a geometric interpretation in mind. The later axiomatic systems for geometry of Peano [4, 5], Hilbert [6], Veblen [7], and Pieri [8] use betweenness or a similar notion of ordered triples as one of the system primitives. More recent discussions of geometry that include betweenness as an undefined notion include Blumenthal [9, 10], Robinson [11] and Tarski [12]. Blank [13] uses primitives of betweenness and a quaternary relation of observed distance to examine the extent to which an individual's visual space is a metric space *.

More directly relevant to the present paper, Huntington and Kline [15] identify eleven sets of independent axioms for strict betweenness that are necessary and sufficient for an agreeing linear order. Their axioms are based on Pasch's and all eleven sets have four axioms in common:

A. $abc \to cba$;

B. $a \neq b \neq c \neq a \to abc$ or acb or bac or bca or cab or cba;

C. $a \neq b \neq c \neq a$ & $abc \to$ not acb;

D. $abc \to a \neq b \neq c \neq a$.

Axioms A and C provide symmetry and a form of one-sided asymmetry. Axioms B and D provide connectedness and strictness (irreflexivity). Various combinations of eight transitivity axioms, two of which are (with a, b, x and y distinct)

1. xab & $aby \to xay$

2. xab & $ayb \to xay$,

are used to complete their eleven sets, one of which is {A, B, C, D, 1, 2}. Van de Walle [16] proves that ten of the eleven Huntington-Kline sets are completely independent in the sense of Moore [17], and Huntington [18] introduces yet another axiom,

9. abc & $x \neq b \to abx$ or xbc,

for a twelfth set {A, B, C, D, 9} of independent axioms.

No axiom of Huntington and Kline uses more than four points. Subsequently, Pitcher and Smiley [19] introduce a series of five-point transitivity axioms. Because they are interested in lattices they use axioms like A and C, but not B because of its connectedness.

* Roberts and Suppes [14] discuss the non-Euclidean nature of, and problems in, the geometry of visual perception.

Another set of axioms that are necessary and sufficient for an agreeing linear order (in the $\leq$ sense rather than the $<$ sense), and which uses a five-point axiom, is given by Sholander [20]. Because of its compactness, I mention it in its entirety. For linear order Sholander uses only three axioms for nonstrict betweenness:

(i) aba iff $a=b$,

(ii) abc & $bde \rightarrow cbd$ or eba,

(iii) abc or bca or cab.

The only disadvantage of this system seems to be that (ii) is not immediately obvious as a necessary condition. Sholander notes one other new axiomatization for linear order, which is much less compact. Roberts [21] also gives a nonstrict betweenness axiomatization for an agreeing linear order ($\leq$)

Sholander's proof for {(i), (ii), (iii)} is based on a nonstrict betweenness theorem of Altwegg [22] whose conditions are similar to the strict betweenness conditions {A, B, C, D, 1, 2} of Huntington and Kline [15]. Altwegg seems to have given the first set of conditions for nonstrict betweenness that are necessary and sufficient for an agreeing partial order $\leq$ (reflexive, transitive, antisymmetric). An alternative system for partial order is given by Sholander, who uses an axiom based on binary comparability, *ab* iff *aab*:

(iv) For odd $n \geq 3$, a_1a_2 & a_2a_3 & ... & $a_{n-1}a_n$ & $a_na_1 \rightarrow$

$$a_{n-1}a_na_1 \text{ or } a_na_1a_2 \text{ or } a_ia_{i+1}a_{i+2} \text{ for some } 1 \leq i \leq n-2.$$

The set {(i), (ii), (iv)} is Sholander's system for an agreeing partial order. Sholander also discusses betweenness in trees and lattices. Axioms to characterize lattices in terms of BT are presented by Smiley and Transue [23].

In another line of development, Huntington [24–26] gives four sets of independent axioms that are necessary and sufficient for his so-called cyclic order, which is a ternary and not a binary relation. Consistent with Huntington's usage we may define BT as a cyclic order iff there is a linear order $<$ such that

$$abc \text{ iff } a<b<c \text{ or } c<a<b \text{ or } b<c<a.$$

A familiar example of a cyclic order arises when X is the perimeter of a circle and *abc* iff the three points are distinct and the clockwise arc from a to b does not include c. One of Huntington's sets of axioms for cyclic order is {B, C, D, E, 9} where the new axiom replacing A (symmetry) is the cyclic property

E. $abc \rightarrow bca$.

Shepperd [27, 28] also touches on cyclic order in discussing "betweenness sets" and "betweenness groups", and Rieger [29] and Fuchs [30, pp. 61–65] examine a notion of cyclically ordered groups using the condition $abc \to (xa, xb, xc) \in \mathrm{BT}$ & $(ay, by, cy) \in \mathrm{BT}$.

Finally, we shall note some cases where betweenness is defined from other concepts.

In connection with a metric space (X, d) a frequent definition of betweenness is abc iff $d(a, b) + d(b, c) = d(a, c)$, with $a \neq b \neq c$ if a strict interpretation is used. See, for example, Kagan [31], Menger [32], Wald [33], Blumenthal [9], Busemann [34] and Krantz [35]. A related definition used by Tversky and Krantz [36], when $X = X_1 \times X_2 \times \ldots \times X_n$, is abc iff $d(a, c) \geq d(a, b)$ & $d(a, c) \geq d(b, c)$ & $a_i = b_i = c_i$ for each $i \in \{1, \ldots, n\}$ for which $a_i = c_i$.

Betweenness has been defined also in terms of qualitative quaternary relations on X, or binary relations on $X \times X$. Beals and Krantz [37], Krantz [35] and Krantz et al. [38] do this in several forms. In connection with $ab \gtrsim cd$ interpreted as "the absolute distance between a and b is at least as large as the absolute distance between c and d", the last of these defines abc iff $ac \gtrsim ab$ & $ac \gtrsim bc$.

Binary relations on X can also be used to define BT. The obvious definition on the basis of a linear order has of course guided some of the axiomatizations mentioned above. Arrow [39] uses this definition in an analysis of the notion of single-peaked preferences of individuals in the theory of social choice *. Birkhoff [41, p. 2] defines BT from a partial order and notes some consequences of the definition.

The following definition ** of lattice betweenness is often used in lattice theory:

L1. $\quad xyz$ iff $(x \cap y) \cup (y \cap z) = y = (x \cup y) \cap (y \cup z)$.

Glivenko [42, 43] observes that xyz by L1 $\to x \cap z \subseteq y \subseteq x \cup z$, and that for metric lattices *** the lattice property of L1 holds iff $d(x, y) + d(y, z) = d(x, z)$, and a metric lattice is distributive if and only if

L2. $\quad xyz$ iff $x \cap z \subseteq y \subseteq x \cup z$,

where $\subseteq$ denotes the partial order for the lattice. Pitcher and Smiley [19] prove more generally that L1 → L2 for any distributive lattice. Smiley [45] examines relationships among lattice betweenness, metric betweenness, and algebraic betweenness (abc iff $b = \lambda a + (1 - \lambda)c$ for some real λ in $[0, 1]$). Restle [46, 47] examines L2 in a context where X is a class of sets.

* Duncan Black's work on single-peaked preferences is the basis for Arrow's analysis. See Black [40].

** Sholander [20, p. 377] mentions another definition of lattice betweenness.

*** See, for example, Glivenko [42], Wilcox and Smiley [44], and Birkhoff [41, p. 76].

Goodman [48] defines a notion of local betweenness that is based on a binary matching (or similarity, etc.) relation. From a psychological viewpoint *, interest in matching stems from the fact that a person will often be unable to detect a "significant" difference between two distinct stimuli. Let $ab \in M$ mean that a and b match or are judged to be "the same". Goodman assumes that X is finite and M is reflexive and symmetric. Let $m(x, y)$ be the number of elements in X that match x or y but not both. Then define LBT (local betweenness) by

$$abc \in \mathrm{LBT} \quad \text{iff} \quad ab \in M,\ bc \in M,\ ac \in M,\ a \neq c\,,$$

$$m(a, c) > m(a, b) \ \text{and} \ m(a, c) > m(b, c),$$

and take *a beside b* iff $a \neq b$ & $ab \in M$ & $acb \in$ LBT for no $c \in X$. Goodman discusses the possibility of M inducing a linear order on X (which occurs iff two distinct elements in X are each beside just one other element and all others are beside exactly two) and notes cases where such an order will not arise. Galanter [50], Coombs [51], Harary [52], and Luce and Galanter [49] discuss this model, and the last of these notes that an induced linear order can give a distorted picture of the underlying stimuli when it is based on a subset of all possible stimuli. Fine [53] provides a lucid mathematical analysis of some of Goodman's ideas.

A more general model than Goodman's linear order model arises in those cases where each element in X can be mapped into a closed real interval with $ab \in M$ iff the intersection of the intervals for a and b is not empty. For finite X, an (X, M) with this property is called an interval graph. We shall consider interval graphs further in the next section.

§2. Orders and interval graphs

The defining properties of the four orders that we shall focus on are:
1. *interval order:* $<$ is irreflexive and $(x < y \ \& \ z < w) \rightarrow (x < w \text{ or } z < y)$, for all $x, y, z, w \in X$;
2. *semiorder:* $<$ is an interval order and $x < y \ \& \ y < z \rightarrow (x < w \text{ or } w < z)$, for all $x, y, z, w \in X$;
3. *weak order:* $<$ is asymmetric and $x < y \rightarrow (x < z \text{ or } z < y)$ for all $x, y, z \in X$;
4. *linear order:* $<$ is irreflexive, transitive, and connected $(x \neq y \rightarrow x < y \text{ or } y < x)$.

Each of these implies its predecessor and can be embedded in its successor. Thus, a semiorder is an interval order, and a semiorder is included in some weak order.

The differences among the orders can be highlighted using the relation $\sim$ defined as follows:

(1) $\qquad x \sim y \ \text{ iff not } \ x < y \ \& \ \text{not } y < x\,.$

* See, for example, Luce and Galanter [49, section 5].

where not $x<y$ means that $(x,y)\notin <$. In all cases $\sim$ is reflexive and symmetric, so that it compares with the matching relation M of the preceding section. This relation has also been called by the names of indifference, similarity, adjacency and incomparability.

If $<$ is a linear order then $x\sim y$ iff $x=y$.

If $<$ is a weak order then $\sim$ is transitive, so that it is an equivalence, and $(\sim)(<)\subseteq <$ and $(<)(\sim)\subseteq <$, where $(\ )(\ )$ denotes composition: $(x,y)\in(\sim)(<)$ iff $x\sim z$ & $z<y$ for some $z\in X$. When $<$ on X is a weak order, $<'$ on $X/\!\sim$ defined by ($r<'s$ iff $x<y$ for some $x\in r$ and $y\in s$) is a linear order.

The term "semiorder" was first used by Luce [54], and the irreflexive definition of semiorder given above is due to Scott and Suppes [55]. If $<$ is a semiorder then $(\sim)(<)\cup(<)(\sim)$ is a weak order (Luce [54, Theorem 1]; Fishburn [56, Theorem 2.8]).

The definition of interval order is from Fishburn [57]. If $<$ is an interval order then $(\sim)(<)$ and $(<)(\sim)$ are weak orders.

Further connections with $\sim$ for interval orders and semiorders come from the notions of interval graphs and indifference graphs. Although my definitions of these concepts differ slightly from other definitions *, the concepts are essentially the same.

An *interval* of a linearly ordered set $(Y,<)$ is a subset I of Y for which $a,b,\in I$ & & $a<c<b\to c\in I$. We shall call $(X,\sim)$ an *interval graph* iff there is a linearly ordered set $(Y,<')$ and a function J on X into the nonempty intervals of $(Y,<')$ such that, for all $x,y\in X$,

$$(2)\qquad x\sim y \text{ iff } J(x)\cap J(y)\neq\emptyset.$$

An *indifference graph* is an interval graph $(X,\sim)$ for which J can be defined so that (2) holds and $J(x)\subset J(y)$ for no $x,y\in X$. The problem of characterizing interval graphs by conditions on $(X,\sim)$ was apparently first proposed by G. Hajós, and independently by Benzer [60] in a study of genetic fine structure in which he examines whether $(X,\sim)$ is an interval graph when X is a set of mutants of a type of gene and $x\sim y$ iff the mutant parts of x and y overlap. Characterizations of interval graphs are given by Lekkerkerker and Boland [61], Gilmore and Hoffman [58], and Fulkerson and Gross [62]. Harary [52] discusses the Gilmore-Hoffman solution along with Goodman's matching theory [48], and Baker et al. [63] note relationships between interval graphs and the Dushnik and Miller [64] definition of the dimension of a partial order.

Roberts [59, 65], who originated the term "indifference graph", presents many interesting aspects of this concept in a finite-set context. Several of his observations [65, Theorems 8 and 9], which draw in part on Fishburn [57], present the connection between interval graphs and orders suggested above.

* See, for example, Gilmore and Hoffman [58] and Roberts [59].

Theorem 0. *$<$ on X is an interval order iff $<$ is transitive and $(X, \sim)$ is an interval graph. $<$ on X is a semiorder iff $<$ is transitive and $(X, \sim)$ is an indifference graph.*

Since the latter part of this theorem has not been proved elsewhere under the definition of indifference graph used here, I shall outline the proof that $(X, \sim)$ is an indifference graph when $<$ on X is a semiorder.

Hence let $<$ on X be a semiorder with $\sim$ defined by (1). Define $x \approx y$ iff $(x \sim z$ iff $y \sim z$, for all $z \in X)$. By Theorem 2.3 in Fishburn [56], $\approx$ is an equivalence and $(x \approx y \;\&\; z \approx w) \to (x < z$ iff $y < w)$. Let Y consist of one element from each class in $X/\approx$. Since $<$ on X is a semiorder it follows that $<$ on Y is a semiorder. In the context of Y let $<' = (\sim)(<) \cup (<)(\sim)$. It follows from Theorems 2.6 and 2.8 in Fishburn [56] that $<'$ on Y is a linear order. Let $\lesssim = (< \cup \sim)$, and for each $x \in Y$ let x^* denote an artificial element that corresponds to x, with Y^* the set of artificial elements (and $Y \cap Y^* = \emptyset$). * Then define $<_o$ on $Y \cup Y^*$ by

$$x <_o y \;\&\; x^* <_o y^* \text{ iff } x <' y$$

$$x^* <_o y \quad \text{iff} \quad x < y$$

$$x <_o y^* \quad \text{iff} \quad x \lesssim y .$$

It is not hard to show that $<_o$ on $Y \cup Y^*$ is a linear order. Finally, for any $a \in X$ with $a \approx x \;\&\; x \in Y$, define $J(a)$ as the closed interval in $(Y \cup Y^*, <_o)$ with x and x^* as end points. By the definition of $<_o$ and the property in the third sentence of this proof it follows easily that $J(a) \cap J(b) \neq \emptyset$ iff $a \sim b$. Moreover, no J interval is properly included in another J interval. For this could be so only if, say, $x <_o y <_o y^* <_o x^*$ for some $x, y \in Y$, and this implies $x <' y$ and $y <' x$ in contradiction of the asymmetry of $<'$. This shows that $(X, \sim)$ is an indifference graph.

If in this proof we take $J(a) <_o J(b)$ to mean that $p <_o q$ whenever $p \in J(a)$ & & $q \in J(b)$, then $a < b$ iff $J(a) <_o J(b)$. Scott and Suppes [55] and Scott [66] show that if $<$ is a semiorder on a finite X then each $J(a)$ can be taken to be a unit interval on the real line.

§3. Agreeing theorems

From the discussion of Section 1 it is clear that many different axiom sets could be used for BT in order to characterize our orders. For each order I have chosen only one set, using the usual criteria of simplicity of expression and interpretability as the guide. Other sets can of course be identified by demonstrating their equivalence to the ones used here.

* That is, Y^* is a copy of Y.

A defined irreflexive and symmetric binary relation of betweenness comparability will be used extensively. It is

$$x/y \quad \text{iff} \quad x \neq y \quad \text{and } x \text{ and } y \text{ both appear in some one triple in BT.}$$

In the usual fashion, $x/y/z/x$ means that x/y & y/z & z/x.

For interval orders and semiorders we shall use the following axioms, which are to hold when indicated for all $x, y, z, w, a, b, c, d \in X$.

A1. $xyz \to zyx$.

A2. $xyz \to$ not yxz.

A3. xyz & $w/y \to xyw$ or wyz and not both.

A4. $x/y/z/x \to xyz$ or zxy or yzx.

A5. abc & $xyz \to abx$ or abz or xyc or zyc.

A6. abc & $xyz \to abx$ or abz or xbc or zbc.

A7. abc & $bcd \to abx$ or xcd or axd.

The first two axioms are similar to A and C of Huntington and Kline [15], and A4 is like their B with $x \neq y \neq z \neq x$ replaced by $x/y/z/x$. A3, which is quite close to Huntington's Axiom 9 of Section 1, expresses the condition that if w/y when xyz then w must lie to one side of y or the other, but not both. Each of A1 through A4 is easily seen to be necessary for an agreeing interval order. Several conditions, some of which were stated earlier, follow from our initial axioms.

Lemma 1. *For all* $x, y, z, w \in X$, $\{\text{A1, A2, A3}\} \to$

C1. *not* xyx & *not* xxy,

C2. $xyz \to$ *not* yxz & *not* yzx,

C3. xyz & $xyw \to$ *not* wyz,

C4. xyz & $yzw \to xyw$ & xzw,

C5. xyz & $xwy \to wyz$ & xwz.

Proof. A2 prohibits xxx. If $x \neq y$, A3 prohibits xyx. A2 prohibits xxy. For C2, use A1 then A2 on zyx to get not yzx. C3 follows from A3: xyz & $y/w \to xyw$ or wyz and not both. Since xyw by hypothesis, not wyz. Using the hypotheses of C4, A3 $\to xyw$ or wyz. But A1 & A2 $\to$ not wyz. Therefore xyw, and wzx follows in like manner. For C5, with xyz & xwy, A3 $\to wyz$ or xyw, the latter of which contradicts C2. Hence wyz. The other conclusion of C5 then follows from C4.

Using C4 and C5, we can build up *strings*, such as $xyzwt$, with the property that when all but three elements are deleted, the remaining three in the given order form a triple in BT. C6 in the next lemma shows how (under A4) another element (x) can be inserted in the middle of a string, and C7 gives another way of extending a string at one end.

Lemma 2. *For all* $a, b, c, d, x, y \in X$, $\{A1, A2, A3, A4\} \to$

C6. abc & bcd & abx & $xcd \to bxc$,

C7. xab & yab & $x/y \to xya$ *or* yxa.

Proof. Use A4 then C3 in each case.

The necessity of A5 for an agreeing interval order is easily verified. Suppose that $<$ is an interval order that agrees with BT. Take abc and xyz as in the hypotheses of A5. For orientation, assume $a < b < c$. If $x < y < z$ then $b < c$ & $y < z \to b < z$ or $y < c$, and hence abz or xyc; if $z < y < x$, a similar calculation gives abx or zyc. Hence one of abz, xyc, abx, and zyc must hold.

The necessity of A6 and A7 for an agreeing semiorder is established in a similar way using the semiorder definition. Neither A6 nor A7 is necessary for an agreeing interval order.

Theorem 1. *There is an interval order on X that agrees with* BT *iff* A1, A2, A3, A4, *and* A5 *hold.*

The sufficiency proof for Theorem 1 is in the next section. Theorems 0 and 1 have the following obvious corollary. $\sim$ is defined by (1).

Corollary 1. *There is a binary relation $<$ on X that agrees with* BT *and is transitive with $(X, \sim)$ an interval graph iff* A1, A2, A3, A4 *and* A5 *hold.*

Theorem 2. *There is a semiorder on X that agrees with* BT *iff* A1, A2, A3, A4, A5, A6 *and* A7 *hold.*

The sufficiency proof for Theorem 2 is in Section 5.

Roberts [21] proves a related theorem for nonstrict betweenness under the assumption that X is finite. His first axiom asserts that (X, I) is an indifference graph when I is defined from nonstrict betweenness by xIy iff xyx. Theorem 2 avoids direct reference to indifference graphs and has, in the presence of Theorem 0, the following corollary.

Corollary 2. *There is a binary relation $<$ on X that agrees with* BT *and is transitive with $(X, \sim)$ an indifference graph iff* A1, A2, A3, A4, A5, A6 *and* A7 *hold.*

The final two theorems add little to previous work, but they do show the modifications in our preceding conditions that lead to agreeing weak orders and linear orders. Two new conditions (to hold for all $x, y, z, w \in X$ when indicated) will be used.

A4*. $x \neq y \neq z \neq x \rightarrow xyz$ or zxy or yzx.

A5*. $xyz \rightarrow wyz$ or xwz or xyw.

A4*, which is essentially Axiom B of Huntington and Kline, will replace A4 in the linear order theorem. A4, as stated ($x/y/z/x \rightarrow xyz$ or zxy or yzx), will be retained in the weak order theorem, where A5* replaces A5. The sufficiency proofs of Theorems 3 and 4 are in Section 6.

Theorem 3. *There is a weak order on X that agrees with* BT *iff* A1, A2, A3, A4 *and* A5* *hold.*

Theorem 4. *There is a linear order on X that agrees with* BT *iff* A1, A2, A3 *and* A4* *hold.*

§4. Proof of Theorem 1

We assume throughout that A1 through A5 hold. If BT = $\emptyset$ then the conclusion is obvious with $< = \emptyset$. Henceforth assume that BT $\neq \emptyset$ and fix abc. On the basis of this fixed triple we define binary $<_1$ in groupings of three thus:

(3) $$x <_1 y <_1 z \quad \text{iff} \quad xyz \ \& \ [(abz \ \& \ \text{not } bzy) \text{ or } (xyc \ \& \ \text{not } bcy)] .$$

This gives $a <_1 b <_1 c$ using A1 and A2. Moreover, by C1 of Lemma 1, $<_1$ is irreflexive. Since $r <_1 s$ and $s <_1 t$ may be based on different triples it is not at all obvious that $<_1$ is transitive. Indeed, the most intricate part of the proof is involved with proving the simpler assertion that $<_1$ is asymmetric.

We note first that $<_1$ agrees with BT. This is followed by several results on $<_1$. Since BT = $\{abc, xyz, abz$, and their three symmetric duals under A1$\}$ satisfies A1–A5 but $<_1$ for this example is not an interval order ($b <_1 c$ & $x <_1 y$, but neither $b <_1 y$ nor $x <_1 c$), we define another binary relation $<_0$ by (4) and then show that the union of $<_0$ and $<_1$ agrees with BT and is indeed an interval order.

Lemma 3. xyz *iff* $x <_1 y <_1 z$ *or* $z <_1 y <_1 x$.

Proof. According to (3), all we need show is that abc & xyz imply one of (abz & not bzy), (xyc & not bcy), (abx & not bxy), and (zyc & not bcy), the last two of

which verify $z <_1 y <_1 x$. By A5 on *abc* & *xyz*, *abz* or *abx* or *xyc* or *zyc*. Suppose that *abz*. If not *bzy* then $x <_1 y <_1 z$. If *bzy* then, using C4, *abzyx*, so that *abx* & *byx*. Since *byx* → not *bxy* by C2 and A1, we get *abx* & not *bxy* so that $z <_1 y <_1 x$. The other three cases (*abx*, *xyc*, *zyc*) are dealt with similarly.

Lemma 4. *xyp & rsp → xsp or ryp or spy.*

Proof. By A1 & A5, *pyx* & *psr* → *pyp* or *pyr* or *psx* or *rsx*. By C1, not *pyp*. Both *pyr* and *psx* are in the conclusion of the lemma under A1. This leaves *rsx*. By C7, (*rsx* & *rsp* & p/x) → *sxp* (which by C4 and C5 gives *rsxyp* and hence *ryp*) or *spx* (which by C5 gives *spy*).

Lemma 5. $<_1$ *is asymmetric.*

Proof. We shall suppose that $x <_1 y$ & $y <_1 x$ and then show that this is impossible. $x <_1 y$ can arise in one of three ways, namely $x <_1 y <_1 t$ or $x <_1 t <_1 y$ or $t <_1 x <_1 y$ for some $t \in X$. These three correspond to the following by (3):

1. *xyt* and [(*abt* & not *bty*) or (*xyc* & not *bcy*)] ,
2. *xty* and [(*aby* & not *byt*) or (*xtc* & not *bct*)] ,
3. *txy* and [(*aby* & not *byx*) or (*txc* & not *bcx*)] .

Similarly, for $y <_1 x$, one of the following must hold for some p:

4. *yxp* and [(*abp* & not *bpx*) or (*yxc* & not *bcx*)] ,
5. *ypx* and [(*abx* & not *bxp*) or (*ypc* & not *bcp*)] ,
6. *pyx* and [(*abx* & not *bxy*) or (*pyc* & not *bcy*)] .

This gives nine major cases to consider (1 & 4, 1 & 5, ..., 3 & 6), and four subcases (from "or") for each major case. Using C7, the subcases which use *ab** for both $i \in \{1, 2, 3\}$ and $j \in \{4, 5, 6\}$ are easily disposed of. For example, for 1 & 6 suppose that (*abt* & not *bty*) and (*abx* & not *bxy*). A1 and C7 then give (since x/t by 1) either *bxt* (which leads by C5 to *bxyt*, contradicting not *bxy*) or *btx* (hence *btyx* by C5, contradicting not *bty*).

By symmetry, it should be clear that we need consider only the following six of the nine main cases: 1 & 4, 1 & 5, 1 & 6, 2 & 5, 2 & 6 and 3 & 6. Having already disposed of the subcases where *ab** holds for i and j, we next dispose of the subcases where ***c* holds for both i and j (for example, where we take (*xyc* & not *bcy*) for 1, and (*yxc* & not *bcx*) for 4). For these subcases, 1 & 4 violates A2; 1 & 5 gives *xypc* by C5, which contradicts *ypx* in 5 by A2; 1 & 6 → not *xyp* by C3, contradicting *pyx* in 6; 2 & 6 → *tcy* by C6 (since *xtyp* & *xtc* & *cyp*), so that *tcb* (contra. not *bct*) or *bcy* (contra. not *bcy*) by A3; 3 & 6 → *xcy* by C6, which by A3 implies *xcb*

(contra. not bcx) or bcy (contra. not bcy). This leaves 2& 5, where it is assumed that xty, (xtc & not bct), ypx, (ypc & not bcp). By Lemma 4, xtc & $ypc \to xpc$ or ytc or pct. First, xpc along with $ypc \to$ not ypx by C3, a contradiction. Second, ytc along with $ytx \to$ not ctx by C3, a contradiction. Finally, pct & $b/c \to bct$ (contra. not bct) or pcb (contra. not bcp).

Twelve subcases remain for the six major cases under consideration. We detail each of these, making use of notational similarity in several cases.

1 & 4. xyt and yxp, hence $pxyt$ by C4.

i. (xyc & not bcy) and (abp & not bpx). Then $pxyc$ so that pyc. Since pba also, Lemma 4 $\to pbc$ (which with $pba \to$ not abc by C3, a contradiction) or bpy (which gives $bpxy$ and hence bpx, a contradiction) or pya. Using C7, pya & pyc & $a/c \to yca$ (hence $ycba$, ycb, contra. not bcy) or yac (hence $pyabc$, pab, contra. abp).

ii. (abt & not bty) and (yxc & not bcx). Similar to (i).

1 & 5. xyt and ypx, hence $xpyt$ by C5.

i. (abt & not bty) and (ypc & not bcp). The interchange of x and p shows that this is similar to 1 & 4 (ii).

ii. (xyc & not bcy) and (abx & not bxp). By interchanging x and p and then noting that pyc and $pxy \to xyc$, we see that this is covered by 1 & 4 (i).

1 & 6. xyt and xyp.

i. (abt & not bty) and (pyc & not bcy). Using Lemma 4, abt & $xyt \to bty$ (contra. not bty) or ayt or xbt.

Consider ayt first. This gives a/y and hence by A3 on xyp, either xya (which with $tya \to$ not xyt, contra.) or ayp. Hence ayp. Then, since cyp & a/c, C7 $\to acy$ (which $\to abcyp \to bcy$, contra.) or cay (which $\to cbayt \to bat$, contra abt).

Finally, suppose that xbt. Then x/b & $abc \to abx$ (which with $tbx \to$ not abt, contra.) or xbc. Hence xbc. Using Lemma 4 on xbc and xyp, get xyc (which with $pyc \to$ not xyp, contra.) or bxy (which $\to bxyt \to bxt$, contra. xbt) or xbp. Hence xbp. Since a/b, A3 $\to xba$ (contra. xbc & abc) or abp. Hence abp. Using Lemma 4 on abp and cyp, get cbp (contra. abp & abc) or bpy ($\to xbpy \to xpy$, contra. xyp) or ayp. Hence ayp. Using C7, ayp & cyp & $a/c \to acy$ ($\to abcyp \to bcy$, contra. not bcy) or cay ($\to cbayp \to bap$, contra. abp).

ii. (xyc & not bcy) and (abx & not bxy). Lemma 4 on abx & xyc gives cbx (contra. abx & abc) or bxy (contra. not bxy) or ayx. Then C7 with $xyc \to acy$ ($\to abcyx$, contra. not bcy) or cay ($\to cbayx$, contra. abx).

2 & 5. xty and xpy.

i. (aby & not byt) and (ypc & not bcp). Lemma 4 on aby and cpy gives apy (which with $cpy \to acp$ ($\to abcpy$, contra. not bcp) or cap ($\to cbapy$, contra. aby)) or cby (contra. aby & abc) or byp ($\to bypx \to byx \to bytx$, contra. not byt).

ii. (xtc & not bct) and (abx & not bxp). Similar to (i).

2 & 6. xty and pyx, hence $xtyp$ by C5.

i. (aby & not byt) and (pyc & not bcy). By A3, pyc & $b/y \to byc$ (with $y/t \to byt$ (contra. not byt) or tyc (contra. pyc & pyt)) or pyb (with pyc by C7 $\to ybc$ (contra. yba & abc) or ycb ($\to pycba$, contra. not bcy)).

ii. (xtc & not bct) and (abx & not bxy). By Lemma 4, abx & $ctx \to atx$ (which with C7 on $ctx \to act$ ($\to abct$, contra. bct) or cat ($\to cbatx$, contra. abx)) or cbx (contra. abx & abc) or bxt ($\to bxty$, contra. not bxy).

3 & 6. txy and pyx, hence $txyp$ by C4.

i. (aby & not byx) & (pyc & not bcy). Interchange x and t, get 2 & 6 (i).

ii. (txc & not bcx) & (abx & not bxy). Similar to (i).

Lemma 6. $x <_1 y <_1 z \ \& \ p <_1 q <_1 r \to p <_1 q <_1 z \ \textit{or} \ x <_1 y <_1 r$.

$$x <_1 y <_1 z \ \& \ p <_1 q <_1 r \to p <_1 y <_1 z \ \textit{or} \ x <_1 q <_1 r.$$

Proof. Assume $x <_1 y <_1 z$ & $p <_1 q <_1 r$. By A5, pqx or pqz or xyr or zyr. By Lemmas 3 and 5, $pqz \to p <_1 q <_1 z$, and $xyr \to x <_1 y <_1 r$. Suppose then that pqx. By Lemmas 3 and 5, $pqx \to p <_1 q <_1 x$. By Lemma 4, pqx & $xyz \to zqx$ (which, since $x <_1 z$, requires $x <_1 q <_1 z$ by Lemmas 3 and 5; but then $x <_1 q$, contradicting $q <_1 x$ by Lemma 5, and hence zqx leads to a contradiction) or pyx (which gives a similar contradiction) or qxy. Hence $pqx \to qxy \to pqxyz \to pqz \to p <_1 q <_1 z$. In like manner, $zyr \to qry \to pqryz \to pqz \to p <_1 q <_1 z$.

The proof of the second half of the lemma is like the proof of the first half.

On the basis of $<_1$, $<_0$ and $<$ are defined as follows.

(4) $\qquad x <_0 y$ iff ($x <_1 r <_1 s$ & $p <_1 q <_1 y$ & not qrs) for some p, q, r, s;

$$< \; = \; <_0 \cup <_1 .$$

Lemma 7a. $x <_1 y \ \& \ z <_1 w \to x < w \ \textit{or} \ z < y$.

Proof. Pairing up one of $x <_1 y <_1 t$, $x <_1 t <_1 y$ and $t <_1 x <_1 y$ with one of $z <_1 w <_1 v$, $z <_1 v <_1 w$ and $v <_1 z <_1 w$, Lemma 6 implies $x <_1 w$ or $z <_1 y$ immediately in all but the following two cases.

1. $x <_1 y <_1 t \ \& \ v <_1 z <_1 w$. If zyt then $z <_1 y$. If not zyt then $x <_0 w$ by (4).
2. $z <_1 w <_1 v \ \& \ t <_1 x <_1 y$. Similar to 1.

Lemma 7b. $x <_0 y \ \& \ z <_0 w \to x < w \ \textit{or} \ z < y$.

Proof. For $x <_0 y$ use (4), and for $z <_0 w$ use $z <_1 j <_1 k \ \& \ m <_1 n <_1 w$ & not njk. By Lemma 6, either $x <_1 j <_1 k$ (whence $x <_0 w$) or $z <_1 r <_1 s$ (whence $z <_0 y$).

Lemma 7c. $x <_0 y$ & $z <_1 w \to x < w$ *or* $z < y$.

Proof. Use (4) for $x <_0 y$. If $z <_1 w <_1 v$ then, by Lemma 6, either $x <_1 w <_1 v$ (so that $x < w$) or $z <_1 r <_1 s$ (so that $z <_0 y$). If $z <_1 v <_1 w$, Lemma 6 $\to x <_1 v <_1 w$ ($x < w$) or $z <_1 r <_1 s$ ($z <_0 y$). If $v <_1 z <_1 w$, Lemma 6 $\to v <_1 z <_1 y$ ($z < y$) or $p <_1 q <_1 w$ ($x <_0 w$).

In summary, Lemmas 7a, 7b and 7c give

Lemma 7. $x < y$ & $z < w \to x < w$ *or* $z < y$.

Lemma 8. $p <_1 q <_1 x$ & $x <_1 r <_1 s \to q <_1 x <_1 r$.

Proof. By Lemma 4 on srx & pqx, either sqx ($\to s <_1 q <_1 x$ by Lemmas 3, 5 and $q <_1 x$, but then $s <_1 x$, contra. $x <_1 r <_1 s$; hence sqx is impossible) or prx (impossible by Lemmas 3 and 5) or qxr ($\to q <_1 x <_1 r$).

Lemma 9. $<$ *is an interval order.*

Proof. In view of Lemma 7 and C1, we need only show that $<_0$ is irreflexive. If $x <_0 x$ then $x <_1 r <_1 s$ & $p <_1 q <_1 x$ & not qrs. But by Lemma 8, $x <_1 r <_1 s$ & $p <_1 q <_1 x \to q <_1 x <_1 r \to qxr \to qxrs \to qrs$. Hence $<_0$ is irreflexive.

Lemma 10. $x < y < z \to x <_1 y <_1 z$.

Proof. Suppose first that $x <_0 y$ and $y <_0 z$ with

$$(5)\qquad \begin{aligned} &x <_1 r <_1 s \ \& \ p <_1 q <_1 y \ \& \ \text{not } qrs, \\ &y <_1 j <_1 k \ \& \ m <_1 n <_1 z \ \& \ \text{not } njk. \end{aligned}$$

By Lemma 8, $q <_1 y <_1 j$. Applying Lemma 6 to this and $x <_1 r <_1 s$, get $q <_1 r <_1 s$ (contra. qrs and $r <_1 s$) or $x <_1 y <_1 j$. Hence $x <_1 y <_1 j$. Also, with Lemma 6 on $q <_1 y <_1 j$ and $m <_1 n <_1 z$, get $q <_1 y <_1 z$ or $m <_1 n <_1 j$. [If $m <_1 n <_1 j$ then with $y <_1 j <_1 k$ Lemma 6 $\to y <_1 j <_1 j$ (false) or $m <_1 n <_1 k$. Then mnk & mnj & $j/k \to nkj$ or njk by C7. If nkj then $j <_1 k <_1 n$ since $j <_1 k$, and thus $j <_1 n$, contra. $m <_1 n <_1 j$. Hence njk. But this contradicts not njk from the hypotheses. Hence $m <_1 n <_1 j$ is false.] Therefore, $q <_1 y <_1 z$. Since x/y from $x <_1 y <_1 j$, A3 $\to xyz$ or qyx (contra. qyj and xyj). Therefore, xyz, and $x <_1 y <_1 z$ since $x <_1 y$.

Suppose next that $x <_1 y$ & $y <_0 z$. We use (5) for $y <_0 z$, and combine this with one of $x <_1 y <_1 t$, $x <_1 t <_1 y$ and $t <_1 x <_1 y$. With $t <_1 x <_1 y$, Lemma 8 $\to x <_1 y <_1 j$. Then Lemma 6 ($m <_1 n <_1 z$) $\to x <_1 y <_1 z$ or $m <_1 n <_1 j$, the latter of which is ruled out by the analysis in brackets in the preceding paragraph.

With $x <_1 t <_1 y$, Lemma 8 $\to t <_1 y <_1 j \to tyj \to xtyj \to xyj \to x <_1 y <_1 j$, and the preceding sentence applies. Thirdly, suppose $x <_1 y <_1 t$. Then Lemma 6 $(y <_1 j <_1 k) \to y <_1 y <_1 t$ (false) or $x <_1 j <_1 k$. By C7, xjk & $yjk \to xyj$ (hence $x <_1 y <_1 j$ and the first use of Lemma 6 in this paragraph applies) or yxj (hence $y <_1 x <_1 j$ since $x <_1 j$; but then $y <_1 x$ contradicts $x <_1 y <_1 t$). Hence $x <_1 y <_1 z$ in all cases.

Finally, suppose that $x <_0 y$ & $y <_1 z$. We use (4) for $x <_0 y$, and one of $y <_1 z <_1 v$, $y <_1 v <_1 z$ and $v <_1 y <_1 z$. Suppose first that $y <_1 v <_1 z$. Then Lemma 8 $\to q <_1 y <_1 v$, and Lemma 6 on this and $x <_1 r <_1 s \to q <_1 r <_1 s$ (contra. not qrs) or $x <_1 y <_1 v$. Hence xyv and thus $xyvz$ and xyz $(\to x <_1 y <_1 z)$. Second, if $y <_1 z <_1 v$ then (Lemma 8) $q <_1 y <_1 z$ and hence $qyzv$ and $q <_1 y <_1 v$ and the two preceding sentences (with $xyvz$ replaced by $xyzv$) apply. Finally, suppose that $v <_1 y <_1 z$. Using Lemma 6 on this and $p <_1 q <_1 y$ get $p <_1 q <_1 z$. Then C7 applied to pqz & pqy & y/z gives qzy or qyz. For qzy, $y <_1 z <_1 q$ since $y <_1 z$. But then $y <_1 q$ contradicts $p <_1 q <_1 y$. Hence not qzy, and therefore qyz $(\to q <_1 y <_1 z)$. Lemma 6 with $x <_1 r <_1 s$ then gives $x <_1 y <_1 z$ or $q <_1 r <_1 s$ (contra. not qrs).

In summary, Lemmas 3, 10 and 7 show that $<$ is an interval order that agrees with BT.

§5. Proof of Theorem 2

Using the two new axioms for an agreeing semiorder,

A6. $\quad abc$ & $xyz \to abx$ or abz or xbc or zbc,

A7. $\quad abc$ & $bcd \to abx$ or xcd or axd,

we shall build from the results of the preceding section. If BT $= \emptyset$ then $< \,= \emptyset$ is an agreeing semiorder. Henceforth assume that BT $\neq \emptyset$. We shall divide the proof into two parts, according to whether a string of four ($abcd$) is formable from BT (using C4 and C5).

PART I. We assume throughout this part that for every $x, y, z, w \in X$, not (xyz & yzw). Since BT $\neq \emptyset$, fix abc, and let this be the abc used to define $<_1$ of the preceding section in (3). Also take $<_0$ defined by (4), and $< \,= \,<_0 \cup <_1$. Furthermore, define

(6) $\quad x <_2 y$ iff $x < p < q$ for some $p, q \in X$ and $y < t$ for no $t \in X$;

$$<' \,= \,< \cup <_2.$$

Lemma 11. $x <' y$ & $y <' z \to x < y < z$.

Proof. Since $x <_2 y \to \text{not } y <' z$, take $x < y$. If $y < z$ then the conclusion follows from Theorem 1. If $y <_2 z$ then we have $x < y$ & $y < r < s$ for some r, s. But this gives a string of four ($xyrs$), contradicting the initial hypothesis of Part I.

Lemma 12. *$<'$ is an interval order.*

Proof. $<$ and $<_2$ are both irreflexive, hence $<'$ is irreflexive. To show that $x <' y \;\&\; z <' w \to x <' w$ or $z <' y$, consider cases:

1. $x < y \;\&\; z < w$. Lemma 7 or 9.

2. $x <_2 y \;\&\; z <_2 w$. Then $x <_2 w \;\&\; y <_2 z$ by (6).

3. $x <_2 y \;\&\; z < w$. Take $x < p < q$ for $x <_2 y$. Then, by Lemma 7, $x < w$ (hence $x <' w$) or $z < p$ (hence $z < p < q$ and thus $z <_2 y$).

Lemma 13. $x <' y \;\&\; y <' z \to x <' w$ *or* $w <' z$.

Proof. By Lemma 11, $x <' y \;\&\; y <' z \to x < y < z$. If $w < t$ for no t then $x <_2 w$ (hence $x <' w$). Henceforth assume that $w < t$ for some t. Then, by (3) and (4) and no string of four, either $w <_1 p <_1 q$ or $p <_1 w <_1 q$ for some $p, q \in X$. Suppose first that $w <_1 p <_1 q$, along with $x <_1 y <_1 z$ (Lemma 10). If $y <_1 p <_1 q$ also then $xypq$ (contra. initial hypothesis), and therefore not ypq. Hence $w <_o z$ by (4), and thus $w <' z$. Suppose next that $p <_1 w <_1 q$. Then by A6 on pwq & xyz, either $p < w < x$ (hence $p < w < x < y < z$, contra. initial hypothesis) or $p < w < z$ (hence $w <' z$) or $x < w < q$ (hence $x <' w$) or $z < w < q$ (hence $x < y < z < w < q$, contra. initial hypothesis).

Together, Lemmas 11, 12 and 13 along with Theorem 1 show that $<'$ is a semiorder that agrees with BT.

PART II. In this part of the proof we assume that abc & bcd, or $abcd$, and let $<$ be any interval order on X that agrees with BT (Theorem 1). With $abcd$ and $<$ fixed with $a < b < c < d$, define

$$x <_3 y \text{ iff not } a < x \;\&\; x < c \;\&\; b < y \;\&\; \text{not } y < d; \tag{7}$$

$$<^* = < \cup <_3.$$

Lemma 14. $x <^* y \;\&\; y <^* z \to x < y < z$.

Proof. $x <_3 y \to b < y \to a < y$ (by transitivity since $<$ is an interval order) $\to$ not $y <_3 z$. Hence, both $<^*$ in the hypotheses of the lemma can't be $<_3$. If $x <_3 y$ & $y < z$ then $a < b < y < z$ and by A7 either $a < b < x$ or $a < x < z$ (each of which

contradicts not $a < x$ from $x <_3 y$) or $x < y < z$. Hence $x < y < z$. Similarly, $x < y \ \& \ y <_3 z \to x < y < z$.

It is easily seen that $<^*$ is irreflexive. The next two lemmas show that $<^*$ is a semiorder which, by Lemma 14 and Theorem 1, agrees with BT.

Lemma 15. $x <^* y \ \& \ y <^* z \to x <^* w$ *or* $w <^* z$.

Proof. By Lemma 14, $x <^* y \ \& \ y <^* z \to x < y < z$. Using $a < b < c < d$ with w, A7 $\to a < b < w$ or $a < w < d$ or $w < c < d$.

1. $a < b < w$. This is similar to the following proof for $w < c < d$.

2. $w < c < d$. By A7 with z, either $a < b < z$ or $a < z < d$ or $z < c < d$. With either of the latter two we get $x < y < z < d$ and, using A7 on this with w, we get $x < w$ or $w < z$ as desired. This leaves $a < b < z$. If $a < w$ then A7 with z applied to $a < w < c < d$ gives $a < w < z$ ($w < z$) or $z < c < d$ (so that $x < y < z < d$ as before) or $a < z < d$ (so that $x < y < z < d$ as before). On the other hand, suppose that not $a < w$. If $z < d$ then $x < y < z < d$ as before, and if not $z < d$ then (not $a < w$ & $w < c$ & $b < z$ & not $z < d$) $\to w <_3 z$.

3. $a < w < d$. A6 on $a < w < d \ \& \ x < y < z \to a < w < x$ (hence $w < z$ under the transitivity of $<$) or $a < w < z$ ($w < z$) or $x < w < d$ ($x < w$) or $z < w < d$ ($x < w$ under transitivity).

Lemma 16. $x <^* y \ \& \ z <^* w \to x <^* w$ *or* $z <^* y$.

Proof. If $x < y \ \& \ z < w$ then $x < w$ or $z < y$ since $<$ is an interval order. If $x <_3 y$ & $z <_3 w$ then $x <_3 w \ \& \ z <_3 y$ by (7). Finally, suppose that $x <_3 y \ \& \ z < w$. By Lemma 7 on $x < c \ \& \ z < w$, either $x < w$ or $z < c$. Suppose that $z < c$. If not $a < z$ then $z <_3 y$ by (7), and if $a < z$ then $a < z < w$ and, by Lemma 15, $a <^* x$ (which is false since $a < x \to$ not $x <_3 y$, and $a <_3 x \to b < x \to a < x$) or $x <^* w$. Hence $x <_3 y \ \& \ z < w \to x <^* w$ or $z <_3 y$.

§6. Proofs of Theorems 3 and 4

Theorem 3 for weak order uses A1–A4 along with A5* ($xyz \to wyz$ or xwz or xyw). To make use of Theorem 1 we first prove

Lemma 17. {A1, A2, A3, A4, A5*} $\to$ A5.

Proof. Given abc & xyz as the hypotheses of A5, use A5* with c on xyz to get cyz or xcz or xyc. Only xcz here is not in the conclusion of A5, which is (abx or abz or xyc or zyc). But xcz & b/c under A3 $\to xcb$ ($\to xcba \to xba$) or bcz($\to abcz \to abz$). Hence $xcz \to abx$ or abz.

Let $<$ be as defined in Section 4. Then, by Lemma 17 and Theorem 1, $<$ is an interval order that agrees with BT. Since an interval order is a weak order when $\sim$, defined by (1), is an equivalence, all we need show is that $\sim$ is transitive. If BT $= \emptyset$ then $< = \emptyset$ and $\sim$ is an equivalence. Henceforth assume that BT $\neq \emptyset$. We then have

Lemma 18. *x/y iff ($x<y$ or $y<x$).*

Proof. Clearly $x/y \to x<y$ or $y<x$. Suppose then that $x<y$. If $x<_1 y$ then x/y. If $x<_0 y$ then, say, $x<r<s$ & $p<q<y$ & not qrs by (4). A5* with y on $x<r<s$ gives $y<r<s$ (which with $p<q<y$ gives $q<r<s$, contradicting not qrs) or $x<y<s$ or $x<r<y$, and therefore x/y.

Thus, by (1), $x \sim y$ iff not ($x<y$ or $y<x$) iff not x/y. If x/z then A5* on each of xzt, xtz and zxt implies either x/y or y/z. Hence not x/y & not $y/z \to$ not x/z, and thus $\sim$ is transitive. This completes the proof of Theorem 3.

Theorem 4 is easily proved either by showing that its axioms imply those of a strict betweenness axiomatization for linear order from Section 1, or by showing that {A1, A2, A3, A4*} $\to$ A4 & A5* and that, when BT $\neq \emptyset$, $x \neq y \to (x<y$ or $y<x)$.

Note added in proof. Lemma 3 says that $xyz \to (x<_1 y<_1 z$ or $z<_1 y<_1 x)$ and $(x<_1 y<_1 z$ or $z<_1 y<_1 x) \to xyz$. Fred Roberts has kindly pointed out that the latter statement has not been proved. Its proof follows from the former statement, A4 and Lemma 5 (whose proof doesn't use Lemma 3): $(x<_1 y<_1 z$ or $z<_1 y<_1 x) \to x/y/z/x \to xyz$ or zxy ($\to z<_1 x<_1 y$ or $y<_1 x<_1 z$, both contra. Lemma 5) or yzx ($\to y<_1 z<_1 x$ or $x<_1 z<_1 y$, both contra. Lemma 5). If $x<_1 y$ is taken to hold iff one of 1, 2, and 3 on p. 169 holds, then (3) emerges as a theorem.

References

[1] H. Freudenthal, The main trends in the foundations of geometry in the 19th century, in: Logic, methodology and philosophy of science, eds. E. Nagel, P. Suppes and A. Tarski (Stanford Univ. Press, Stanford, California, 1962).

[2] M. Pasch, Vorlesungen über neuere Geometrie, 1st. ed. (Teubner, Leipzig, 1882).

[3] M. Pasch, Grundlagen der Geometrie, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Vol. 23 (Springer, Berlin, 1926).

[4] G. Peano, Sui fondamenti della geometria, Rivista di Matematica 4 (1894) 51–90.

[5] G. Peano, I principii di geometria (Turin, 1889).

[6] D. Hilbert, Grundlagen der Geometrie (1st ed. 1899; 7th ed. Teubner, Leipzig and Berlin, 1930).

[7] O. Veblen, A system of axioms for geometry, Trans. Amer. Math. Soc. 5 (1904) 343–384.

[8] M. Pieri, La geometria elementare instituita sulle nozioni di 'punto' e 'sfera', Memorie di Matematica e de Fisica della Societa Italiana delle Scienze, Serie III, 15 (1908) 345–450.

[9] L.M. Blumenthal, Distance geometries: a study of the development of abstract metrics, The Univ. of Missouri Studies 13 (1938).
[10] L.M. Blumenthal, A modern view of geometry (Freeman, San Francisco, 1961).
[11] G. de B. Robinson, The foundations of geometry (Univ. of Toronto Press, Toronto, Canada, 1940).
[12] A. Tarski, What is elementary geometry ?, pp. 16–29, in: Symposium on the axiomatic method, eds. L. Henkin, P. Suppes and A. Tarski (North-Holland, Amsterdam, 1959).
[13] A.A. Blank, Axiomatics of binocular vision: the foundations of metric geometry in relation to space perception, J. Opt. Soc. Am. 48 (1958) 328–333.
[14] F.S. Roberts and P. Suppes, Some problems in the geometry of visual perception, Synthese 17 (1967) 173–201.
[15] E.V. Huntington and J.R. Kline, Sets of independent postulates for betweenness, Trans. Am. Math. Soc. 18 (1917) 301–325.
[16] W.E. van de Walle, On the complete independence of the postulates for betweenness, Trans. Am. Math. Soc. 26 (1924) 249–256.
[17] E.H. Moore, Introduction to a form of general analysis, New Haven colloquium 1906 (Yale Univ. Press, New Haven, 1910).
[18] E.V. Huntington, A new set of postulates for betweenness, with proof of complete independence, Trans. Am. Math. Soc. 26 (1924) 257–282.
[19] E. Pitcher and M.F. Smiley, Transitivities of betweenness, Trans. Am. Math. Soc. 52 (1942) 95–114.
[20] M. Sholander, Trees, lattices, order and betweenness, Proc. Am. Math. Soc. 3 (1952) 369–381.
[21] F.S. Roberts, Tolerance geometry, RAND paper P-4430, The RAND Corporation, Santa Monica, California, 1970.
[22] M. Altwegg, Zur Axiomatik der teilweise geordneten Mengen, Commen. Math. Helv. 24 (1950) 149–155.
[23] M.F. Smiley and W.R. Transue, Applications of transitivities of betweenness in lattice theory, Bull. Am. Math. Soc. 49 (1953) 280–287.
[24] E.V. Huntington, A set of independent postulates for cyclic order, Proc. Natl. Acad. Sci. U.S. 2 (1916) 630–631.
[25] E.V. Huntington, Sets of completely independent postulates for cyclic order, Proc. Natl. Acad. Sci. U.S. 10 (1924) 74–78.
[26] E.V. Huntington, Inter-relations among the four principal types of order, Trans. Am. Math. Soc. 38 (1935) 1–9.
[27] J.A.H. Shepperd, Transitivities of betweenness and separation and the definitions of betweenness and separation groups, J. London Math. Soc. 31 (1956) 240–248.
[28] J.A.H. Shepperd, Betweenness groups, J. London Math. Soc. 32 (1957) 277–285.
[29] L.S. Rieger, On the ordered and cyclically ordered groups, I–III, Věstník Král. České Spol. Nauk. no. 6 (1946) 1–31; no. 1 (1947) 1–33; no. 1 (1948) 1–26 (in Czech.).
[30] L. Fuchs, Partially ordered algebraic systems (Addison-Wesley, Reading, Mass., 1963).
[31] B. Kagan, Ein System von Postulaten welche die Euklidische Geometrie definieren, Jahresber. Deutschen Math. Ver. 11 (1902) 403–424.
[32] K. Menger, Untersuchungen über allgemeine Metrik, Math. Ann. 100 (1928) 75–163.
[33] A. Wald, Axiomatik des Zwischenbegriffes in metrischen Räumen, Math. Ann. 104 (1931) 476–484.
[34] H. Busemann, The geometry of geodesics (New York, 1955).
[35] D.H. Krantz, A survey of measurement theory, pp. 314–350 in: Mathematics of the decision sciences, part 2, eds. G.B. Dantzig and A.F. Veinott Jr. (Amer. Math. Soc., Providence, Rhode Island, 1968).

[36] A. Tversky and D.H. Krantz, The dimensional representation and the metric structure of similarity data, Michigan Math. Psychol. Program MMPP 69–7, Univ. of Michigan, Ann Arbor, 1969.
[37] R. Beals and D.H. Krantz, Metrics and geodesics induced by order relations, Math. Z. 101 (1967) 285–298.
[38] D.H. Krantz, R.D. Luce, P. Suppes and A. Tversky, Foundations of measurement, Vol. 1 (Academic Press, New York, 1971, in press).
[39] K.J. Arrow, Social choice and individual values, 2nd ed. (Wiley, New York, 1963).
[40] D. Black, The theory of committees and elections (Cambridge Univ. Press, Cambridge, England, 1958).
[41] G. Birkhoff, Lattice theory, rev. ed. (Amer. Math. Soc., New York, 1948).
[42] V. Glivenko, Géométrie des systèmes de choses normées, Am. J. Math. 58 (1936) 799–828.
[43] V. Glivenko, Contribution à l'étude des systèmes de choses normées, Am. J. Math. 59 (1937) 941–956.
[44] L.R. Wilcox and M.F. Smiley, Metric lattices, Ann. Math. 40 (1939) 309–327.
[45] M.F. Smiley, A comparison of algebraic, metric, and lattice betweenness, Bull. Amer. Math. Soc. 49 (1943) 246–252.
[46] F. Restle, A metric and an ordering on sets, Psychometrika 24 (1959) 207–220.
[47] F. Restle, Psychology of judgment and choice: a theoretical essay (Wiley, New York, 1961).
[48] N. Goodman, The structure of appearance (Harvard Univ. Press, Cambridge, Mass., 1951).
[49] R.D. Luce and E. Galanter, Discrimination, Chapter 4 in: Handbook of mathematical psychology, vol. 1, eds. R.D. Luce, R.R. Bush and E. Galanter (Wiley, New York, 1963).
[50] E. Galanter, An axiomatic and experimental study of sensory order and measure, Psychol. Rev. 63 (1956) 16–28.
[51] C.H. Coombs, A theory of data (Wiley, New York, 1964).
[52] F. Harary, A graph theoretic approach to similarity relations, Psychometrika 29 (1964) 143–151.
[53] N.J. Fine, Proof of a conjecture of Goodman, J. Symbolic Logic 19 (1954) 41–44.
[54] R.D. Luce, Semiorders and a theory of utility discrimination, Econometrica 24 (1956) 178–191.
[55] D. Scott and P. Suppes, Foundational aspects of theories of measurement, J. Symbolic Logic 23 (1958) 113–128.
[56] P.C. Fishburn, Utility theory for decision making (Wiley, New York, 1970).
[57] P.C. Fishburn, Intransitive indifference with unequal indifference intervals, J. Math. Psychol. 7 (1970) 144–149.
[58] P.C. Gilmore and A.J. Hoffman, A characterization of comparability graphs and of interval graphs, Can. J. Math. 16 (1964) 539–548.
[59] F.S. Roberts, Indifference graphs, pp. 139–146 in: Proof techniques in graph theory, ed. F. Harary (Academic Press, New York, 1969).
[60] S. Benzer, On the topology of the genetic fine structure, Proc. Natl. Acad. Sci. U.S. 45 (1959) 1607–1620.
[61] C.G. Lekkerkerker and J. Ch. Boland, Representation of a finite graph by a set of intervals on the real line, Fund. Math. 51 (1962) 45–64.
[62] D.R. Fulkerson and O.A. Gross, Incidence matrices and interval graphs, Pacific J. Math. 15 (1965) 835–855.
[63] K.A. Baker, P.C. Fishburn and F.S. Roberts, Partial orders of dimension 2, interval orders, and interval graphs, RAND paper P-4376, The RAND Corporation, Santa Monica, California, 1970.
[64] B. Dushnik and E.W. Miller, Partially ordered sets, Am. J. Math. 63 (1941) 600–610.
[65] F.S. Roberts, On nontransitive indifference, J. Math. Psychol. 7 (1970) 243–258.
[66] D. Scott, Measurement structures and linear inequalities, J. Math. Psychol. 1 (1964) 233–247.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 2 (1971) pp. 179–197.

LIMITS IN CATEGORIES OF RELATIONS AND LIMIT–COLIMIT COMMUTATION

Armin FREI and John L. MACDONALD
University of British Columbia, Vancouver, BC, USA
Forschungsinstitut für Mathematik, ETH Zürich, Switzerland

Received 14 December 1970

The category $\tilde{\mathscr{A}}$ of relations in an Abelian category $\mathscr{A}$ is isomorphic to its own dual. This entails that direct limits in $\tilde{\mathscr{A}}$ can be computed as inverse limits and vice-versa. This, together with the fact that limits of the same type commute, suggests the use of the category $\tilde{\mathscr{A}}$ to obtain criteria for limit–colimit commutation in $\mathscr{A}$.

The first section consists of an account of the necessary introductory paraphernalia. The second section is devoted to cofinal functors and their applications to the theory of limits and relative limits. In the third section conditions are given for the existence of certain limits and relative limits in the category $\tilde{\mathscr{A}}$ of relations. In section four we show that in $\tilde{\mathscr{A}}$ limits can be computed as colimits and vice-versa. Under certain circumstances the limit of the colimit functor of a bifunctor F with range $\mathscr{A}$ is just the limit in $\tilde{\mathscr{A}}$ of a suitable functor. A key point for applications is given by a theorem giving conditions for the commutativity of limits and relative limits in $\tilde{\mathscr{A}}$. In section five we apply the foregoing to the problem of commuting of limits with colimits in $\mathscr{A}$.

§ 1. Paraphernalia

The terminology used will generally be that of [2] and [3]. Thus, to a commutative square

$$(1.1) \qquad \begin{array}{ccc} R & \xrightarrow{\alpha} & A \\ \beta\downarrow & & \downarrow\varphi \\ B & \xrightarrow[\psi]{} & X \end{array}$$

in an Abelian category $\mathscr{A}$ we associate the differential sequence

$$(1.2)\qquad R \xrightarrow{\{\alpha,\beta\}} A \oplus B \xrightarrow{\langle\varphi,\,-\psi\rangle} X\,.$$

Whenever (1.2) is exact, left exact, right exact we say that (1.1) is exact, cartesian, cocartesian and write ex$(\alpha,\beta;\varphi,\psi)$, car$(\alpha,\beta;\varphi,\psi)$, coc$(\alpha,\beta;\varphi,\psi)$, respectively. Clearly car$(\alpha,\beta;\varphi,\psi)$ means (1.1) is a pullback diagram and coc$(\alpha,\beta;\varphi,\psi)$ that (1.1) is a pushout.

The category $\tilde{\mathscr{A}}$ has the same objects as $\mathscr{A}$. A morphism from A to B in $\tilde{\mathscr{A}}$ is an equivalence class of morphism pairs (φ,ψ) in $\mathscr{A}$ where

$$A \xrightarrow{\varphi} X \xleftarrow{\psi} B$$

and X is an object in $\mathscr{A}$. Two morphism pairs (φ,ψ) and (φ',ψ') are equivalent if and only if there are monomorphisms μ and μ' such that $\mu\varphi = \mu'\varphi'$ and $\mu\psi = \mu'\psi'$. We write φ/ψ for the class of (φ,ψ) and note that every such class has a minimal representative, unique up to isomorphism in $\mathscr{A}$, given by the pushout of the pullback of any representative (φ,ψ). Conversely any pushout is a minimal representative. The composition $\chi/\lambda \circ \varphi/\psi$ is given by $\chi'\varphi/\psi'\lambda$ where (χ',ψ') is any pair for which the square in

$$\begin{tikzcd}
{} \arrow[r, "\varphi"] & {} \arrow[r, "\chi'"] & {} \\
& {} \arrow[u, "\psi"] \arrow[r, "\chi"'] & {} \arrow[u, "\psi'"'] \\
& & {} \arrow[u, "\lambda"']
\end{tikzcd}$$

is exact.

We identify $\mathscr{A}$ with a subcategory of $\tilde{\mathscr{A}}$ via the embedding functor $\varphi \to \varphi/1$. If $F\colon \mathfrak{C} \to \mathscr{A}$ is a functor, then $\tilde{F}\colon \mathfrak{C} \to \tilde{\mathscr{A}}$ will denote its composition with the embedding functor. The isomorphisms of $\tilde{\mathscr{A}}$ are just those of $\mathscr{A}$. An anti-involution $\tilde{T}\colon \tilde{\mathscr{A}} \to \tilde{\mathscr{A}}$ is defined on $\tilde{\mathscr{A}}$ by $\tilde{T}(\varphi/\psi) = \psi/\varphi$. We write $\overline{\varphi/\psi}$ short for $\tilde{T}(\varphi/\psi)$ and note that $\varphi/\psi = \bar{\psi} \circ \varphi$.

By a process dual to that used for constructing $\tilde{\mathscr{A}}$ a category $\tilde{\tilde{\mathscr{A}}}$ is constructed. There is an isomorphism $\tilde{\mathscr{A}}$ to $\tilde{\tilde{\mathscr{A}}}$ which carries the morphism φ/ψ of $\tilde{\mathscr{A}}$ to $\alpha\backslash\beta\colon A \xleftarrow{\alpha} R \xrightarrow{\beta} B$ of $\tilde{\tilde{\mathscr{A}}}$, where (α,β) is any pair for which (1.1) is exact. Notice that a commutative square (1.1) is exact if and only if $\beta\bar{\alpha} = \bar{\psi}\varphi$. As before we denote by $\tilde{\tilde{F}}\colon \mathfrak{C} \to \tilde{\tilde{\mathscr{A}}}$ the composition of $F\colon \mathfrak{C} \to \mathscr{A}$ with the embedding functor.

We say that a morphism φ/ψ of $\tilde{\mathscr{A}}$ is represented minimally when it is minimal in its class and similarly we speak of maximal representation in $\tilde{\tilde{\mathscr{A}}}$.

We now turn to a few propositions to be used in the sequel.

Proposition 1.3: *Suppose that in the diagram*

$$(1.4)\qquad \begin{tikzcd} & \cdot \arrow[dl, "\psi"'] \arrow[dr, "\psi'"] & \\ \cdot \arrow[rr, dashed, "\alpha'"] & & \cdot \\ \cdot \arrow[u, "\varphi"] \arrow[rr, "\alpha"'] & & \cdot \arrow[u, "\varphi'"'] \end{tikzcd}$$

(φ, ψ) *is minimal and* $\varphi/\psi = \varphi'/\psi' \circ \alpha$. *Then* $\exists|$ *a monic* α' *with* $\alpha'\varphi = \varphi'\alpha$ *and* $\alpha'\psi = \psi'$.

Proposition 1.5. *Suppose that in the diagram* (1.4) (φ, ψ) *is minimal and* $\alpha \circ \psi/\varphi =$ $= \psi'/\varphi'$. *Then* $\exists|$ α' *such that* $\alpha'\varphi = \varphi'\alpha$ *and* $\alpha'\psi = \psi'$. *Furthermore* $\mathrm{ex}(\alpha, \varphi; \varphi', \alpha')$. *If in addition* (ψ', φ') *is minimal, then* $\mathrm{coc}(\alpha, \varphi; \varphi', \alpha')$.

Proof. Let $\mathrm{coc}(\alpha, \varphi; \varphi'', \alpha')$. Then $\alpha'\psi/\varphi'' = \psi'/\varphi'$ and (ψ, φ), being minimal, is a pushout, hence $(\varphi'', \alpha'\psi)$ is a pushout. Hence $(\alpha'\psi, \varphi'')$ is the minimal representative of ψ'/φ', thus $\exists|$ a monic θ with

$$(1.6)\qquad \theta\alpha'\psi = \psi'$$

and $\theta\varphi'' = \varphi'$, hence

$$(1.7)\qquad \theta\alpha'\varphi = \varphi'\alpha\,.$$

As (ψ, φ) is a pushout it follows that $\theta\alpha'$ is the unique morphism for which (1.6) and (1.7) hold. From θ monic and $\mathrm{coc}(\alpha, \varphi; \varphi'', \alpha')$ we infer $\mathrm{ex}(\alpha, \varphi; \varphi', \theta\alpha')$. If (ψ', φ') is minimal, then θ is an isomorphism.

From [2] we have

Theorem 1.8. *Consider the commutative diagram*

$$\begin{tikzcd} \cdot \arrow[r, "\varphi"] & \cdot & \\ \cdot \arrow[u, "\alpha"] \arrow[r, "\beta"'] \arrow[drr, "\beta'"'] & \cdot \arrow[u, "\psi"'] \arrow[dr, "\omega"] & \\ & & \cdot \arrow[uul, "\psi'"'] \end{tikzcd}$$

in $\mathcal{A}$. *Then* $\mathrm{coc}(\alpha, \beta; \varphi, \psi)$ *and* $\mathrm{ex}(\alpha, \beta'; \varphi, \psi')$ *implies* ω *epic;* $\mathrm{ex}(\alpha, \beta; \varphi, \psi)$ *and* $\mathrm{car}(\alpha, \beta'; \varphi, \psi')$ *implies* ω *monic;* $\mathrm{coc}(\alpha, \beta; \varphi, \psi)$ *and* $\mathrm{car}(\alpha, \beta'; \varphi, \psi')$ *implies* ω *is an isomorphism.*

We conclude this section with

Proposition 1.9. *If in the commutative diagram*

$$\begin{tikzcd} \cdot \arrow[r, "\alpha"] \arrow[d, "\beta"'] & \cdot \arrow[r, "\alpha'"] \arrow[d, "\varphi"] & \cdot \arrow[d, "\theta"] \\ \cdot \arrow[r, "\psi"'] & \cdot \arrow[r, "\psi'"'] & \cdot \end{tikzcd}$$

(i) $\mathrm{car}(\alpha'\alpha, \beta; \theta, \psi'\psi)$ *and* $\mathrm{car}(\alpha', \varphi; \theta, \psi')$, *then* $\mathrm{car}(\alpha, \beta; \varphi, \psi)$. (ii) $\mathrm{ex}(\alpha'\alpha, \beta; \theta, \psi'\psi)$ *and* $\mathrm{car}(\alpha', \varphi; \theta, \psi')$, *then* $\mathrm{ex}(\alpha, \beta; \varphi, \psi)$.

§2. Cofinal functors

In the sequel we shall assume that all index categories I, J are small, nonempty and connected. J is said to be *quasi-filtered* (write qf) if in addition

Given $\begin{tikzcd}[row sep=tiny] & \cdot \\ \cdot \arrow[ur, "\alpha"] \arrow[dr, "\beta"'] & \\ & \cdot \end{tikzcd}$ in J there are γ, δ with $\gamma\alpha = \delta\beta$.

Given $\overset{\rho}{\underset{\sigma}{\rightrightarrows}}$ in J there is η with $\eta\rho = \eta\sigma$.

A category I is called qf′ if it satisfies the dual of the preceding conditions.

Let J, J' be any categories. A functor $C\colon J' \to J$ is said to be *cofinal* if it satisfies the following conditions:

(i) Given $j \in |J|$ there is a morphism $j \to Cj'$ for some $j' \in |J'|$.

(ii) Given

$$\begin{tikzcd}[row sep=small] & Cj'_1 \\ j \arrow[ur, "\alpha_1"] \arrow[dr, "\alpha_2"'] & \\ & Cj'_2 \end{tikzcd}$$

in J, there are morphisms ψ_1, ψ_2 in J' such that

$$\begin{tikzcd}[row sep=small] & Cj'_1 \arrow[dr, "C\psi_1"] & \\ j \arrow[ur, "\alpha_1"] \arrow[dr, "\alpha_2"'] & & Cj'_3 \\ & Cj'_2 \arrow[ur, "C\psi_2"'] & \end{tikzcd} \tag{2.1}$$

commutes. One verifies that whenever (i) holds, (ii) is equivalent to either of the following, provided J and J' are qf:

(ii)′ Given

$$j \overset{\varphi_1}{\underset{\varphi_2}{\rightrightarrows}} Cj'$$

in J there is a morphism ψ in J' for which $C\psi \circ \varphi_1 = C\psi \circ \varphi_2$.

(ii)″ Given

$$Cj'_1 \overset{\varphi}{\underset{C\sigma}{\rightrightarrows}} Cj'_2$$

in J, there is a morphism ψ in J' for which $C\psi \circ \varphi = C\psi \circ C\sigma$.

(We owe this notion of cofinal functor to Peter Hilton.)

In particular let J' be a full subcategory of the qf category J and suppose that J' satisfies the following condition:

Given $j \in |J|$ there is a morphism $j \to j'$ in J with $j' \in |J'|$. Then J' is qf and the inclusion functor is a cofinal functor.

As another example let J be a qf category and J' the comma category j/J of objects under j. Then j/J is qf and the projection functor P: $j/J \to J$ is cofinal.

Let $F: J \to \mathfrak{C}$ be any functor. A natural transformation $\kappa: F \to X$, where X is a constant functor (i.e. $X \in |\mathfrak{C}|$) will be called a *cone over F*. Similarly a natural transformation $\eta: X \to F$ will be called a *cone under F*.

From now on we will write "Suppose (C, J', J)" for "Suppose that $C: J' \to J$ is cofinal".

Lemma 2.2. *Suppose (C, J', J). Let $F: J \to \mathfrak{C}$ be any functor, then every cone over FC can be extended in a unique way to a cone over F.*

Proof. Let $\rho_{Ci}: FCi \to X$, $i \in |J'|$ be a cone. Let $j \in |J|$. Then there exists a $\varphi: j \to Ci$ and we set

$$\rho_j = \rho_{Ci} \circ F\varphi . \tag{2.3}$$

Let $\varphi': j \to Ci'$. Then by (ii), definition of cofinal functor, there are ψ, ψ' in J' such that

$$\begin{tikzcd}[row sep=small] & Ci \arrow[dr, "C\psi"] & \\ j \arrow[ur, "\varphi"] \arrow[dr, "\varphi'"'] & & Ck \\ & Ci' \arrow[ur, "C\psi'"'] & \end{tikzcd}$$

commutes. Hence $\rho_j = \rho_{Ci} \circ F\varphi = \rho_{Ck} \circ FC\psi \circ F\varphi = \rho_{Ck} \circ FC\psi' \circ F\varphi' = \rho_{Ci'} \circ F\varphi'$. This shows that ρ_j, as defined, is independent of the choice of φ, furthermore the ρ_j's agree with the given ρ_{Ci}'s for any j of the form $j = Ci$, $i \in |J'|$. Let now $\psi: j \to k$ be in J. Then $\exists\theta: k \to Ci$, $i \in |J'|$. By (2.3) $\rho_{Ci} \circ F\theta \circ F\psi = \rho_j$ and $\rho_{Ci} \circ F\theta = \rho_k$. Hence $\rho_j = \rho_k \circ F\psi$. Thus ρ is a cone over F. ρ is clearly unique since any extension must satisfy equation (2.3).

Clearly every cone π over F can be restricted to a cone π_C over FC, hence under the hypotheses of Lemma 2.2 there is a 1–1 correspondence between cones over F and cones over FC.

Given a class $\mathcal{F}$ of cones over F we shall denote $\mathcal{F}_C$ the class of the corresponding cones over FC.

Definition 2.4. Let $F: J \to \mathfrak{C}$ be any functor and $\mathcal{F}$ a class of cones over F. A cone $\pi: F \to L$ in $\mathcal{F}$ is said to be a *direct $\mathcal{F}$-limit* of F (we write $\{L, \pi\} = \mathcal{F}\text{-}\varinjlim F$) if, given any cone $\rho: F \to X$ in $\mathcal{F}$, there exists a unique $\mathfrak{C}$-morphism $\gamma: L \to X$ such that $\gamma\pi = \rho$.

Clearly $\mathcal{F}\text{-}\varinjlim F$ is unique up to canonical isomorphism.

Theorem 2.5. *Suppose (C, J', J). Let $\mathcal{F}$ be a class of cones over $F: J \to \mathfrak{C}$. Then $\{L, \pi\} = \mathcal{F}\text{-}\varinjlim F \leftrightarrow \{L, \pi_C\} = \mathcal{F}_C\text{-}\varinjlim FC$.*

Proof. Let $\{L, \pi_C\} = \mathcal{F}_C\text{-}\varinjlim FC$ and $\rho: F \to X$ a cone in $\mathcal{F}$. Then $\exists!$ $\mathfrak{C}$-morphism $\sigma: L \to X$ with $\sigma \circ \pi_C = \rho_C$. But $\forall j \in |J|$ there is a morphism $\psi: j \to Ci$ in J and

$$\rho_j = \rho_{Ci} \circ F\psi = \sigma \circ \pi_{Ci} \circ F\psi = \sigma\pi_j .$$

Conversely, let $\{L, \pi\} = \mathcal{F}\text{-}\varinjlim F$ and $\rho_C: FC \to X$ a cone in $\mathcal{F}_C$. Then $\exists!$ $\mathfrak{C}$-morphism $\sigma: L \to X$ with $\rho = \sigma\pi$. Clearly $\rho_C = \sigma \circ \pi_C$ and if $\rho_C = \sigma'\pi_C$, then as before $\rho = \sigma'\pi$. Hence $\sigma' = \sigma$ and this shows the uniqueness of σ.

If in this last theorem we let $\mathcal{F}$ be the class of all cones over F, we obtain

Theorem 2.6. *Suppose (C, J', J). Then $\{L, \pi\} = \varinjlim F \leftrightarrow \{L, \pi_C\} = \varinjlim FC$.*

As a consequence of Theorem 2.5 we have the following result.

Theorem 2.7. *Suppose (C, J', J). Let $F: J \to \mathfrak{C}$ and $G: \mathfrak{C} \to \mathfrak{D}$ be any functors. Furthermore let $\mathcal{F}_1$, $\mathcal{F}_2$ be any classes of cones over F, GF respectively. Then $G(\mathcal{F}_1\text{-}\varinjlim F) = \mathcal{F}_2\text{-}\varinjlim GF$ if and only if $G((\mathcal{F}_1)_C\text{-}\varinjlim FC) = (\mathcal{F}_2)_C\text{-}\varinjlim GFC$. In particular, $G(\varinjlim F) = \varinjlim GF$ if and only if $G(\varinjlim FC) = \varinjlim GFC$.*

Remark. A functor between qf categories which is cofinal in the sense of Bass ([1], Definition 8.3, page 46) is cofinal in our sense.

Let $C: J' \to J$ be any functor and $\kappa: G \to F: J \to \mathfrak{C}$ a natural transformation. For each $\psi: j_1 \to j$ in J' we have a commutative diagram

(2.8)
$$\begin{tikzcd}
GCj_1 \arrow[r, "GC\psi"] \arrow[d, "\kappa_{Cj_1}"'] & GCj \arrow[d, "\kappa_{Cj}"] \\
FCj_1 \arrow[r, "FC\psi"'] & FCj
\end{tikzcd}$$

Let $\{LG, \lambda\} = \varinjlim G$ and $\{LF, \pi\} = \varinjlim F$. Then, for every $j \in |J'|$ we have the commutative diagram

$$(2.9)\qquad \begin{tikzcd} GCj \arrow[r, "\lambda_{Cj}"] \arrow[d, "\kappa_{Cj}"'] & LG \arrow[d, "L\kappa"] \\ FCj \arrow[r, "\pi_{Cj}"'] & LF \end{tikzcd}$$

For this situation we prove

Theorem 2.10. *Suppose (C, J', J) and let j_1 be initial object in J'. If (2.8) is cocartesian for all $\psi: j_1 \to j$, then (2.9) is cocartesian for every $j \in |J'|$.*

Proof. We only prove that (2.9) is cocartesian for $j = j_1$. By the dual of Proposition 1.9, (i), it follows readily that (2.9) is cocartesian for every $j \in |J'|$.

We consider the object FCj_1 as a constant functor $J' \to \mathfrak{C}$. There is a natural transformation $\mu: FCj_1 \to FC$ given by $\mu_j = FC\psi$, where ψ is the unique morphism $j_1 \to j$. Since C is cofinal $\{LF, \pi_C\} = \varinjlim FC$ and $\pi_{Cj_1} = \pi_{Cj} \circ FC\psi$, $\forall j \in |J'|$. But $\varinjlim_{J'} \mu = \pi_{Cj} \circ FC\psi$, $\forall j \in |J'|$, thus $\pi_{Cj_1} = \varinjlim_{J'} \mu$. We consider the commutative diagram of functors $J' \to \mathfrak{C}$ and natural transformations

$$\begin{tikzcd} GCj_1 \arrow[r, "\eta_j"] \arrow[d, "\kappa_{Cj_1}"'] & GCj \arrow[d, "\kappa_{Cj}"] \\ FCj_1 \arrow[r, "\mu_j"'] & FCj \end{tikzcd}$$

where η plays the same role for G as μ does for F. By hypothesis the above diagram is cocartesian for every $j \in |J'|$. By taking direct limits we obtain the diagram (2.9), for $j = j_1$, which is cocartesian since direct limits preserve cocartesian squares.

If in Theorem 2.10 we let C be the projection functor $P: j_1/J \to J$ for any $j_1 \in |J|$, then $1_{j_1} \in |j_1/J|$ is initial and we obtain

Corollary 2.11: *Let J be a* qf *category. If*

$$\begin{tikzcd} Gj_1 \arrow[r, "G\psi"] \arrow[d, "\kappa_{j_1}"'] & Gj \arrow[d, "\kappa_j"] \\ Fj_1 \arrow[r, "F\psi"'] & Fj \end{tikzcd}$$

is cocartesian for fixed $j_1 \in |J|$ and for all $\psi: j_1 \to j$, then

$$\begin{tikzcd} Gj \arrow[r, "\lambda_j"] \arrow[d, "\kappa_j"'] & LG \arrow[d, "L\kappa"] \\ Fj \arrow[r, "\pi_j"'] & LF \end{tikzcd}$$

is cocartesian for every j of the form $P\psi$.

Proof. Notice that if J is qf, then $P: j_1/J \to J$ is cofinal.

§3. Limits and relative limits in $\tilde{\mathscr{A}}$

We shall denote by capital Roman letters $A, B, X, \ldots$ objects of the Abelian category $\mathscr{A}$ and by lower case Greek letters $\alpha, \beta, \chi, \ldots$ morphisms of $\mathscr{A}$ whether they are considered in $\mathscr{A}$ or $\tilde{\mathscr{A}}$. Throughout this paper we assume the existence of all limits in $\mathscr{A}$ which arise.

Definition 3.1. A functor $F: J \to \mathscr{A}$ is said to be d-conservative if $\varinjlim \tilde{F} = \varinjlim F$, similarly F is said to be i-conservative if $\varprojlim \tilde{F} = \varprojlim F$. Clearly $\tilde{F}$ can be replaced by $\tilde{\tilde{F}}: J \to \tilde{\tilde{\mathscr{A}}}$.

Note that under the canonical isomorphism $\hat{\mathscr{A}}$ and $\tilde{\tilde{\mathscr{A}}}$ can always be identified. Let $F: J \to \mathscr{A}$ be a functor and X in $|\mathscr{A}|$ i.e. X a constant functor $J \to \mathscr{A}$. Let

$$\alpha/\beta = \{ Fj \xrightarrow{\alpha_j} Gj \xleftarrow{\beta_j} X \}: \tilde{F} \to X$$

be a cone over $\tilde{F}$. If α/β is given minimally, then the Gj's are uniquely determined up to canonical isomorphism. Furthermore, from (1.3) it follows that $\forall \psi: j \to k$ in J there is a unique monomorphism $G\psi$ which renders

(3.2)
$$\begin{tikzcd} & X \arrow[dl, "\beta_j"'] \arrow[dr, "\beta_k"] & \\ Gj \arrow[rr, "G\psi"] & & Gk \\ Fj \arrow[u, "\alpha_j"] \arrow[rr, "F\psi"'] & & Fk \arrow[u, "\alpha_k"'] \end{tikzcd}$$

commutative. It is easy to verify that G is indeed a functor $J \to \mathscr{A}$. Let $\{LF, \pi\} = \varinjlim F$ and $\{LG, \lambda\} = \varinjlim G$. Then for every $i \in |J|$ we obtain the commutative

diagram

$$(3.3)\qquad \begin{tikzcd} & X \arrow[dl, "\beta_i"'] \arrow[dr, "L\beta"] & \\ Gi \arrow[rr, "\lambda_i"] & & LG \\ Fi \arrow[u, "\alpha_i"] \arrow[rr, "\pi_i"'] & & LF \arrow[u, "L\alpha"'] \end{tikzcd}$$

We are now ready to prove

Theorem 3.4. *If J is* qf *and $\mathscr{A}$ satisfies the Grothendieck axiom* AB5, *then every functor $F: J \to \mathscr{A}$ is* d-*conservative.*

Proof. The morphisms λ_i of (3.3) are monics since the $G\psi$'s of (3.2) are monics, J is qf and $\mathscr{A}$ is AB5. This follows from [5], 14. 6.4 Hilfssatz. Hence $L\alpha/L\beta \circ \pi_i = \alpha_i/\beta_i$. Let $LF \xrightarrow{\alpha'} Y \xleftarrow{\beta'} X$ represent a morphism in $\tilde{\mathscr{A}}$ for which $\alpha'/\beta' \circ \pi_i = \alpha_i/\beta_i$, $\forall i \in |J|$. We show that $\alpha'/\beta' = L\alpha/L\beta$. By Proposition 1.3 $\exists|\,\mathscr{A}$-monic $\epsilon_i: Gi \to Y$ with $\epsilon_i\beta_i = \beta'$ and $\epsilon_i\alpha_i = \alpha'\pi_i$ for every $i \in |J|$. $\epsilon: G \to Y$ is clearly a cone over G in $\mathscr{A}$, hence there exists a unique $\mathscr{A}$-morphism $L\epsilon: LG \to Y$ with $L\epsilon \circ \lambda_i = \epsilon_i$, $\forall i \in |J|$. $L\epsilon$ is monic since $\mathscr{A}$ is AB5 and J is qf. Furthermore

$$L\epsilon \circ L\beta = L\epsilon \circ \lambda_i \circ \beta_i = \epsilon_i\beta_i = \beta'$$

and

$$L\epsilon \circ L\alpha \circ \pi_i = L\epsilon \circ \lambda_i \circ \alpha_i = \epsilon_i\alpha_i = \alpha'\pi_i\,, \qquad \forall i \in |J|\,,$$

thus $L\epsilon \circ L\alpha = \alpha'$. Hence $L\alpha/L\beta = \alpha'/\beta'$.

We leave the formulation of the dual of Theorem 3.4 to the reader. However, we point out that if an Abelian category $\mathscr{A}$ satisfies both AB5 and its dual, then $\mathscr{A}$ consists of zero objects only.

As before a cone over $\tilde{F}: J \to \tilde{\mathscr{A}}$ in $\tilde{\mathscr{A}}$ is a natural transformation $\alpha\backslash\beta$ in $\tilde{\mathscr{A}}$ from $\tilde{F}$ to a constant functor $X: J \to \tilde{\mathscr{A}}$. The cone $\alpha\backslash\beta$ is called *cocartesian* if it has at least one cocartesian representation, that is, a representation $\{Fj \xleftarrow{\alpha_j} Gj \xrightarrow{\beta_j} X\}$ for which there exists a functor $G: J \to \mathscr{A}$ such that the diagram

$$(3.5)\qquad \begin{tikzcd} & X & \\ Gj \arrow[ur, "\beta_j"] \arrow[rr, "G\psi"] \arrow[d, "\alpha_j"'] & & Gk \arrow[ul, "\beta_k"'] \arrow[d, "\alpha_k"] \\ Fj \arrow[rr, "F\psi"'] & & Fk \end{tikzcd}$$

commutes in $\mathscr{A}$ for every $\psi: j \to k$ in J, and such that the square in (3.5) is cocartesian.

Remark. A cone $\alpha\backslash\beta$ is cocartesian if and only if its maximal representation is cocartesian.

In a dual fashion we define cartesian cones under $\tilde{F}: I \to \tilde{\mathscr{A}}$.

Suppose (C, J', J). If $F: J \to \mathscr{A}$ is a functor, let $\mathcal{F}$ be the class of all cones $\alpha\backslash\beta$ over $\tilde{\tilde{F}}$ for which $(\alpha\backslash\beta)_C$ is cocartesian. Then $\mathcal{F}_C$ is the class of all cocartesian cones over $\tilde{\tilde{F}}C$.

Theorem 3.6. *Suppose (C, J', J) and that J' is* qf. *Let $F: J \to \mathscr{A}$ be a functor and $\{LF, \pi\} = \varinjlim F$. Then*

$$\{LF, \pi_C\} = \mathcal{F}_C\text{-}\varinjlim \tilde{\tilde{F}}C \tag{3.7}$$

and equivalently

$$\{LF, \pi\} = \mathcal{F}\text{-}\varinjlim \tilde{\tilde{F}}\,. \tag{3.8}$$

Proof. If $\alpha\backslash\beta$ is a cone over $\tilde{\tilde{F}}$ belonging to $\mathcal{F}$, then by hypothesis $(\alpha\backslash\beta)_C: \tilde{\tilde{F}}C \leftarrow G \to X$ is a cocartesian cone over $\tilde{\tilde{F}}C$ with associated functor $G: J' \to \mathscr{A}$. Let $\{LG, \lambda\} = \varinjlim G$. In the commutative diagram

$$\begin{tikzcd} & X & \\ Gi \arrow[ur, "\beta_{Ci}"] \arrow[rr, "\lambda_i"] \arrow[d, "\alpha_{Ci}"'] & & LG \arrow[ul, "L\beta"'] \arrow[d, "L\alpha"] \\ FCi \arrow[rr, "\pi_{Ci}"'] & & LF \end{tikzcd} \tag{3.9}$$

the square is cocartesian by Corollary 2.11. Hence $L\alpha\backslash L\beta \circ \pi_{Ci} = \alpha_{Ci}\backslash\beta_{Ci}$, $\forall i \in |J'|$. Suppose $\mu\backslash\eta \circ \pi_{Ci} = \alpha_{Ci}\backslash\beta_{Ci}$, $\forall i \in |J'|$, where (μ, η) is maximal. Then by Proposition 1.5 (dual) there exists, $\forall i \in |J'|$, a unique $\mathscr{A}$-map $\epsilon_i: Gi \to Y$ such that the diagram

$$\begin{tikzcd} & X & \\ Gi \arrow[ur, "\beta_{Ci}"] \arrow[rr, "\epsilon_i"] \arrow[d, "\alpha_{Ci}"'] & & Y \arrow[ul, "\eta"'] \arrow[d, "\mu"] \\ FCi \arrow[rr, "\pi_{Ci}"'] & & LF \end{tikzcd} \tag{3.10}$$

commutes and has exact square. Clearly $\epsilon: G \to Y$ is a cone over G, hence there exists a unique $\mathscr{A}$-map $L\epsilon: LG \to Y$ with

$$(3.11) \qquad L\epsilon \circ \lambda_i = \epsilon_i \,.$$

Furthermore $L\epsilon$ satisfies

$$(3.12) \qquad \mu \circ L\epsilon = L\alpha$$

and

$$(3.13) \qquad \eta \circ L\epsilon = L\beta \,.$$

From (3.11), (3.12) and Theorem 1.8 we infer that $L\epsilon$ is epic, which together with (3.12) and (3.13) entails $\mu\backslash\eta = L\alpha\backslash L\beta$. Hence (3.7) holds.

By Theorem 2.5 it follows that (3.7) is equivalent to (3.8).

Corollary 3.14. *Suppose (C, J', J) and J'* qf. *If $F: J \to \mathscr{A}$ is a functor for which every cone over $\tilde{\tilde{F}}C$ is cocartesian, then F is* d-*conservative.*

Proof. As before, let $\mathscr{F}$ be the class of cones over $\tilde{\tilde{F}}$ having $\mathscr{F}_C$ equal to the class of all cocartesian cones over $\tilde{\tilde{F}}C$. By hypothesis $\mathscr{F}_C$ is the class of all cones over $\tilde{\tilde{F}}C$, hence by Lemma 2.2 $\mathscr{F}$ is the class of all cones over $\tilde{\tilde{F}}$.

We apply the preceding to obtain

Theorem 3.15. *Suppose (C, J', J) and J'* qf. *If $F: J \to \mathscr{A}$ is a functor for which $FC\psi$ is epic for all morphisms ψ in J', then F is* d-*conservative.*

Proof. Let $\alpha\backslash\beta$: $FC \leftarrow G \to X$ be a cone given maximally. Then by Proposition 1.5 (dual) for each $\psi: j \to k$ in J' there exists a unique $\mathscr{A}$-map $G\psi: Gj \to Gk$ such that (3.5) commutes and has cartesian square. But by hypothesis $F\psi$ in (3.5) is epic, hence the square is also cocartesian. This shows that every cone over $\tilde{\tilde{F}}C$ is cocartesian.

In a previous example we have seen that if J is qf, then the projection functor $P: j/J \to J$ is cofinal. Hence if J is qf, and $F: J \to \mathscr{A}$ has the property that for fixed j and variable $\psi: j \to k$ in J, $F\psi$ is epic, then F is d-conservative.

§4. Bifunctors and limits in $\tilde{\mathscr{A}}$

In this section we first point out that $\tilde{\mathscr{A}}$ is isomorphic to $\tilde{\mathscr{A}}^{\circ}$. As a consequence we have that when the direct limit of a functor $H: J \to \tilde{\mathscr{A}}$ exists it can be computed as the inverse limit of the contravariant functor obtained by composing H with the anti-involution of $\tilde{\mathscr{A}}$, and vice-versa. More precisely, we have

Proposition 4.1. *Let $\widetilde{T}: \widetilde{\mathcal{A}} \to \widetilde{\mathcal{A}}$ be the anti-involution and $H: J \to \widetilde{\mathcal{A}}$ a functor for which $\varinjlim H$ exists. If $\{L, \pi\} = \varinjlim_{J} H$, then $\{L, \bar{\pi}\} = \varprojlim_{J^o} \widetilde{T}H$.*

Let $F: I \times J \to \mathcal{A}$ be a functor, and $\widetilde{F}$ the composition of F with the embedding $\mathcal{A} \to \widetilde{\mathcal{A}}$. For every $\varphi: i \to i_1$ in I and $\psi: j_1 \to j$ in J we obtain a commutative diagram

$$(4.2)\qquad \begin{array}{ccc} F(i, j_1) & \xrightarrow{F(i,\psi)} & F(i,j) \\ {\scriptstyle F(\varphi, j_1)}\downarrow & & \downarrow{\scriptstyle F(\varphi, j)} \\ F(i_1, j_1) & \xrightarrow[F(i_1, \psi)]{} & F(i_1, j) \end{array}$$

To the functor $\widetilde{F}$ we make to correspond the two mappings

$\widetilde{F}_{\ell}: I^o \times J \to \widetilde{\mathcal{A}}$ defined by $\widetilde{F}_{\ell}(i,j) = F(i,j)$,

$\widetilde{F}_{\ell}(\varphi^o, \psi) = F(i_1, \psi)/F(\varphi, j) = \overline{F(\varphi, j)} \circ F(i_1, \psi)$ and similarly

$\widetilde{F}_{r}: I \times J^o \to \widetilde{\mathcal{A}}$ defined by $\widetilde{F}_{r}(i,j) = F(i,j)$,

$\widetilde{F}_{r}(\varphi, \psi^o) = F(\varphi, j)/F(i_1, \psi) = \overline{F(i_1, \psi)} \circ F(\varphi, j)$.

Clearly $\widetilde{F}_{\ell} = \widetilde{T}\widetilde{F}_{r}$.

Proposition 4.3. *$\widetilde{F}_{\ell}$ is a functor $\leftrightarrow$ (4.2) is exact $\leftrightarrow$ $\widetilde{F}_{r}$ is a functor.*

Proof. We prove only the first equivalence.

(4.2) is exact $\leftrightarrow \overline{F(\varphi, j)} \circ F(i_1, \psi) = F(i, \psi) \circ \overline{F(\varphi, j_1)}$

$\leftrightarrow \widetilde{F}_{\ell}(\varphi, j) \circ \widetilde{F}_{\ell}(i_1, \psi) = \widetilde{F}_{\ell}(i, \psi) \circ \widetilde{F}_{\ell}(\varphi, j_1)$

$\leftrightarrow \widetilde{F}_{\ell}$ is a functor.

Let $F: I \times J \to \mathcal{A}$ be a functor and $Fi: J \to \mathcal{A}$ the corresponding functor for fixed $i \in |I|$. From now on we shall denote by $\{LFi, \pi_{ij}\}$ the direct limit of Fi and by $\{RFj, \tau_{ij}\}$ the inverse limit of Fj. Given a morphism $\varphi: i \to i_1$ in I the diagram

$$(4.4)\qquad \begin{array}{ccc} F(i,j) & \xrightarrow{\pi_{ij}} & LFi \\ {\scriptstyle F(\varphi, j)}\downarrow & & \downarrow{\scriptstyle LF\varphi} \\ F(i_1, j) & \xrightarrow[\pi_{i_1 j}]{} & LFi_1 \end{array}$$

obviously commutes $\forall j \in |J|$. But

$$(4.5)\qquad \begin{tikzcd} F(i,j) \arrow[r, "\pi_{ij}"] & LFi \\ F(i_1,j) \arrow[u, "\overline{F(\varphi,j)}"] \arrow[r, "\pi_{i_1 j}"'] & LFi_1 \arrow[u, "\overline{LF\varphi}"'] \end{tikzcd}$$

commutes if and only if (4.4) is exact. If (4.2) is exact for every (φ, ψ) in $I \times J$, then $\pi_{ij} \circ \overline{F(\varphi, j)}$ is a cone over $\tilde{F}i_1$. Thus, if in addition Fi_1 is d-conservative and (4.4) is exact for all $j \in |J|$, then $\overline{LF\varphi}$ is the only map which renders (4.5) commutative since $\{LFi_1, \pi_{ij}\} = \varinjlim_J \tilde{F}i_1$.

Theorem 4.6. *Let I, J be any categories. Let $F: I \times J \to \mathscr{A}$ be a functor for which (4.2) is exact for every (φ, ψ) in $I \times J$ and (4.4) is exact for every (φ, j) in $I \times J$. Suppose in addition that for every $i \in |I|$ the functor $Fi: J \to \mathscr{A}$ is d-conservative and that the functor $\varinjlim_J F: I \to \mathscr{A}$ is i-conservative. If $\{LF, \pi_{ij}\} = \varinjlim_J F$ and $\{RLF, \tau_i\} = \varprojlim_I LF$, then $\{RLF, \pi_{ij}/\tau_i\} = \varinjlim_{I^{\circ} \times J} \widetilde{F}_{\varrho}$.*

Proof. We first show that π_{ij}/τ_i is a cone over $\widetilde{F}_{\varrho}$. For $\psi: j_1 \to j$ in J we have $\bar{\tau}_i \circ \pi_{ij} \circ F(i, \psi) = \bar{\tau}_i \circ \pi_{ij_1}$ and by the definition of $\widetilde{F}_{\varrho}$ we have $\pi_{ij}/\tau_i \circ \widetilde{F}_{\varrho}(i, \psi) = \pi_{ij_1}/\tau_i$. For $\varphi: i \to i_1$ in I we have

$$\begin{aligned} \pi_{i_1 j}/\tau_{i_1} &= \overline{LF\varphi \circ \tau_i} \circ \pi_{i_1 j} \\ &= \bar{\tau}_i \circ \overline{LF\varphi} \circ \pi_{i_1 j} \\ &= \bar{\tau}_i \circ \pi_{ij} \circ \overline{F(\varphi, j)} \text{ since (4.4) is exact,} \\ &= \pi_{ij}/\tau_i \circ \widetilde{F}_{\varrho}(\varphi, j) \ \text{ by definition of } \widetilde{F}_{\varrho}. \end{aligned}$$

Now let $\alpha/\beta: \widetilde{F}_{\varrho} \to\!\!\leftarrow X$ be a cone over $\widetilde{F}_{\varrho}$. Since Fi is d-conservative $\{LFi, \pi_{ij}\} = \varinjlim_J \widetilde{F}i = \varinjlim_J (\widetilde{F}_{\varrho})i$, hence $\exists\,|$ map $\mu_i/\eta_i: LFi \to X$ in $\widetilde{\mathscr{A}}$ with $\mu_i/\eta_i \circ \pi_{ij} = \alpha_{ij}/\beta_{ij}$.

Furthermore, for any $\varphi: i \to i_1$ in I, $\mu_i/\eta_i \circ \overline{LF\varphi} \circ \pi_{i_1 j} = \mu_{i_1}/\eta_{i_1} \circ \pi_{i_1 j}$, hence $\mu_i/\eta_i \circ \overline{LF\varphi} = \mu_{i_1}/\eta_{i_1}$, i.e. $LF\varphi \circ \eta_i/\mu_i = \eta_{i_1}/\mu_{i_1}$. But LF is i-conservative, hence $\exists|\ \sigma/\vartheta: X \to\!\!\leftarrow RLF$ with $\tau_i \circ \sigma/\vartheta = \eta_i/\mu_i$. Hence $\vartheta/\sigma \circ \bar{\tau}_i \circ \pi_{ij} = \mu_i/\eta_i \circ \pi_{ij}$ and $\vartheta/\sigma \circ \pi_{ij}/\tau_i = \alpha_{ij}/\beta_{ij}$. It is straightforward that ϑ/σ is the unique morphism with this last property.

Given the functor $F: I \times J \to \mathscr{A}$ let $\{RF, \tau_{ij}\} = \varprojlim_I F$ and $\{LRF, \pi_j\} = \varinjlim_J RF$.

From Theorem 4.6 and its dual we infer

Theorem 4.7. *Let I, J be any categories. Let $F: I \times J \to \mathcal{A}$ be a functor for which* (4.2), (4.4) *and*

$$
\begin{tikzcd}
RFj_1 \arrow[r, "RF\psi"] \arrow[d, "\tau_{ij_1}"'] & RFj \arrow[d, "\tau_{ij}"] \\
F(i,j_1) \arrow[r, "F(i,\psi)"'] & F(i,j)
\end{tikzcd}
\tag{4.8}
$$

are exact for all (φ, ψ), (φ, j) *and* (i, ψ) *in* $I \times J$ *respectively. Suppose in addition that*

(i) *For all* $i \in |I|$, $Fi: J \to \mathcal{A}$ *is d-conservative and* LF *is* i-conservative.

(ii) *For all* $j \in |J|$, $Fj: I \to \mathcal{A}$ *is* i-*conservative and* RF *is* d-*conservative.*

Then $\{RLF, \pi_{ij}/\tau_i\} = \varinjlim_{I^o \times J} \widetilde{F}_\varrho$ *and* $\{LRF, \tau_{ij}\backslash\pi_j\} = \varprojlim_{I \times J^o} \widetilde{F}_\mathrm{r}$. *Furthermore,* $RLF \cong LRF$,

i.e. by abuse of language $\varprojlim_I \varinjlim_J F \cong \varinjlim_J \varprojlim_I F$.

Notice that $RLF \cong LRF$ follows from $\widetilde{T}\widetilde{F}_\varrho = \widetilde{F}_\mathrm{r}$ and Proposition 4.1.

Let the functor $D: I' \to I$ be coinitial, with I, I' qf' and suppose (C, J', J). Then $D^o: I'^o \to I^o$ is cofinal and $D^o \times C: I'^o \times J' \to I^o \times J$ is cofinal, with $I'^o \times J'$ and $I^o \times J$ qf.

If $\widetilde{F}_\varrho$ is a functor, then $\varinjlim \widetilde{F}_\varrho \cong \varinjlim \widetilde{F}_\varrho(D^o \times C)$ when either side is defined. The previous theorem applied to the functor $[\widetilde{F(D \times C)}]_\varrho = \widetilde{F}_\varrho(D^o \times C)$ gives a criterion for $\varinjlim \widetilde{F}_\varrho(D^o \times C)$ to exist.

As an application of Theorem 4.7 we obtain

Theorem 4.9. *Let I be* qf' *and J* qf. *Let $F: I \times J \to \mathcal{A}$ be a functor for which* (4.2) *is exact for every* (φ, ψ) *in* $I \times J$, $F(\varphi, j)$ *is monic for all* (φ, j) *in* $I \times J$, *and* $F(i, \psi)$ *is epic for every* (i, ψ) *in* $I \times J$. *Then the conclusion of Theorem* 4.7 *holds.*

Proof. (4.2) is clearly bicartesian and by Corollary 2.11 and its dual, (4.4) and (4.8) are cocartesian and cartesian respectively which, in particular, entails that $RF\psi$ is epic for every ψ in J and $LF\varphi$ is monic for every $\varphi \in I$. Hence by Theorem 3.15, RF and Fi $(\forall i \in |I|)$ are d-conservative and by the dual of Theorem 3.15, LF and Fj $(\forall j \in |J|)$ are i-conservative.

If $F(\varphi, j_1)$ is monic, $F(i_1, \psi)$ is epic and (4.2) is exact for fixed $(i_1, j_1) \in |I \times J|$ and variable $\varphi: i \to i_1$ and $\psi: j_1 \to j$, then the conclusion of Theorem 4.7 still holds. Indeed if $Q: I/i_1 \to I$ and $P: j_1/J \to J$ are the projection functors, then $\varprojlim_I \varinjlim_J \widetilde{F}(Q \times P) \cong$

$\cong \varprojlim_I \varinjlim_J \widetilde{F}$ and $\varinjlim_J \varprojlim_I \widetilde{F}(Q \times P) \cong \varinjlim_J \varprojlim_I \widetilde{F}$. Furthermore, using Proposition 1.9 one shows that $F(Q \times P)$ satisfies the hypotheses of Theorem 4.9. (A detailed treatment of a similar situation is given in the proof of Corollary 4.15.)

In order to obtain a limit commutation theorem which is applicable to a wider range of functors $F: I \times J \to \mathscr{A}$ we make incisive use of the notion of relative limit. If $H: I \to \mathscr{A}$ is any functor we shall denote by $\mathcal{F}\text{-}\varprojlim \widetilde{H}$ the *cartesian* cone under $\widetilde{H}$ which is universal with respect to all cartesian cones under $\widetilde{H}$.

Theorem 4.10. *Let I be* qf$'$. *Let $F: I \times J \to \mathscr{A}$ be a functor for which $Fi: J \to \mathscr{A}$ is* d-*conservative for every $i \in |I|$. Suppose in addition that* (4.4) *is cartesian for every* (φ, j) *in $I \times J$, and that*

$$(4.11) \qquad \begin{array}{ccc} RFj & \xrightarrow{\pi_j} & LRF \\ {\scriptstyle \tau_{ij}} \downarrow & & \downarrow {\scriptstyle L\tau_i} \\ F(i,j) & \xrightarrow[\pi_{ij}]{} & LFi \end{array}$$

is exact for every (i, j) in $|I \times J|$.

Then $LRF \cong RLF$, i.e. *by abuse of language* $\varinjlim_J \varprojlim_I F \cong \varprojlim_I \varinjlim_J F$.

Remark. If (4.4) is cartesian, then by Proposition 1.9 (4.2), is cartesian. If (4.2) is cartesian, then by Corollary 2.11 (dual), (4.8) is cartesian. If we assume that direct limits preserve the cartesianity of (4.2) and (4.8) then (4.4) and (4.11) are cartesian.

Proof. Since (4.11) is exact

$$(4.12) \qquad \begin{array}{ccc} RFj & \xleftarrow{\bar{\pi}_j} & LRF \\ {\scriptstyle \tau_{ij}} \downarrow & & \downarrow {\scriptstyle L\tau_i} \\ F(i,j) & \xleftarrow[\bar{\pi}_{ij}]{} & LFi \end{array}$$

commutes in $\widetilde{\mathscr{A}}$. The morphism $\bar{\pi}_j$ is the unique one in $\widetilde{\mathscr{A}}$ for which (4.12) commutes. Indeed $L\tau_i/\bar{\pi}_{ij}: LRF \to LFi \leftarrow F(i,j)$ is a cartesian cone since (4.4) is cartesian. Hence $\exists!\ \vartheta_j: LRF \to RFj$ in $\widetilde{\mathscr{A}}$ with $\tau_{ij} \circ \vartheta_j = L\tau_i/\pi_{ij}$. Thus $\vartheta_j = \bar{\pi}_j$.

From (4.4) we obtain

$$(4.13) \qquad \begin{array}{ccc} RFj & \xrightarrow{R\pi_j} & RLF \\ {\scriptstyle \tau_{ij}} \downarrow & & \downarrow {\scriptstyle \tau_i} \\ F(i,j) & \xrightarrow[\pi_{ij}]{} & LFi \end{array}$$

by taking inverse limits. Since (4.4) is cartesian so is (4.13) by Corollary 2.11 (dual), hence

(4.14)
$$\begin{tikzcd} RFj \arrow[d, "\tau_{ij}"'] & RLF \arrow[l, dashed, "\overline{R\pi_j}"'] \arrow[d, "\tau_i"] \\ F(i,j) & LFi \arrow[l, "\bar\pi_{ij}"] \end{tikzcd}$$

commutes in $\tilde{\mathscr{A}}$. τ_i/π_{ij}: $RLF \to LFi \leftarrow F(i,j)$ is a cartesian cone since (4.4) is cartesian. Hence, by the same argument as before, $\overline{R\pi_j}$ is the unique map in $\tilde{\mathscr{A}}$ rendering (4.14) commutative. $R\pi_j$: $RFj \to RLF$ is a cone (in $\mathscr{A}$) over RF. Hence $\exists\,|\ \rho$: $LRF \to RLF$ in $\mathscr{A}$ with $\rho\pi_j = R\pi_j$. Thus $\bar\pi_j \circ \bar\rho = \overline{R\pi_j}$, $\forall j \in |J|$.

$L\tau_i$. $LRF \to LFi$ is a cone (in $\mathscr{A}$) under LF. Hence $\exists\,|\ \omega$: $LRF \to RLF$ in $\mathscr{A}$ with $\tau_i \circ \omega = L\tau_i$. It is also unique in $\tilde{\mathscr{A}}$ with this property since all cones in $\mathscr{A}$ are cartesian.

We now show that $\bar\rho\omega = 1_{LRF}$. We have

$$\begin{aligned} \tau_{ij} \circ \bar\pi_j \circ \bar\rho \circ \omega &= \tau_{ij} \circ \overline{R\pi_j} \circ \omega \\ &= \bar\pi_{ij} \circ \tau_i \circ \omega \quad \text{since (4.14) commutes} \\ &= \bar\pi_{ij} \circ L\tau_i = \tau_{ij} \circ \bar\pi_j \quad \text{since (4.12) commutes.} \end{aligned}$$

Hence $\bar\pi_j \circ \bar\rho \circ \omega = \bar\pi_j$, $\forall j \in |J|$, since both sides render (4.12) commutative, and $\bar\pi_j$ is the unique morphism with this property. But $\bar\pi_j \circ \bar\rho \circ \omega = \bar\pi_j$ holds if and only if $\bar\omega \circ \rho \circ \pi_j = \pi_j$ does. Now π_j: $RFj \to LRF$ is a cocartesian cone, hence by Theorem 3.6 there is a unique morphism $\nu\ (= 1_{LRF})$ in $\tilde{\mathscr{A}}$ for which $\nu \circ \pi_j = \pi_j$, $\forall j \in |J|$. Hence $\bar\omega \circ \rho = 1_{LRF}$ and clearly $\bar\rho \circ \omega = 1_{LRF}$.

We now show that $\omega \circ \bar\rho = 1_{RLF}$. We have

$$\begin{aligned} \bar\pi_{ij} \circ \tau_i \circ \omega \circ \bar\rho &= \bar\pi_{ij} \circ L\tau_i \circ \bar\rho \\ &= \tau_{ij} \circ \bar\pi_j \circ \bar\rho \quad \text{since (4.12) commutes} \\ &= \tau_{ij} \circ \overline{R\pi_j} = \bar\pi_{ij} \circ \tau_i \quad \text{since (4.14) commutes.} \end{aligned}$$

Since $\{LFi, \pi_{ij}\} = \varinjlim_J \widetilde{F}i$ (since Fi is d-conservative) the preceding implies that $\bar\tau_i = \rho \circ \bar\omega \circ \bar\tau_i$, i.e. that $\tau_i = \tau_i \circ \omega \circ \bar\rho$. But this implies that $\omega \circ \bar\rho = 1_{RLF}$ since τ_i: $RLF \to LFi$, being a cone in $\mathscr{A}$, is cartesian. Thus LRF is isomorphic to RLF in $\tilde{\mathscr{A}}$ and hence in $\mathscr{A}$. This completes the proof.

By using cofinality we can slightly weaken the hypotheses of Theorem 4.10. For this let $Q\colon I/i_1 \to I$ and $P\colon j_1/J \to J$ be the projection functors. Then we have

Corollary 4.15. *Let I be* qf$'$ *and J* qf. *Let $(i_1, j_1) \in |I\times J|$ be fixed. Let $F\colon I\times J \to \mathcal{A}$ be a functor for which Fi is* d-*conservative for every $i = Q(i \to i_1)$. Suppose in addition that* (4.2) *is cartesian for every $\varphi\colon i \to i_1$ and every $\psi\colon j' \to j$ with $\psi = P\psi$,* (4.4) *is cartesian for every $\varphi\colon i \to i_1$ and every $j = P(j_1 \to j)$, and finally that* (4.11) *is exact for $i = i_1$ and every $j = P(j_1 \to j)$. Then $LRF \cong RLF$.*

Proof. *If $F(Q\times P)\colon I/i_1 \times j_1/J \to \mathcal{A}$* satisfies the hypotheses of Theorem 4.10, then $LRF \cong RLF$. Indeed, by Theorem 4.10 $\varinjlim_{j_1/J} \varprojlim_{I/i_1} F(Q\times P) \cong \varprojlim_{I/i_1} \varinjlim_{j_1/J} F(Q\times P)$. But $F\colon J \to \mathcal{A}^I$, therefore, by the cofinality of P, $\varinjlim_J F \cong \varinjlim_{j_1/J} F(1\times P)\colon I \to \mathcal{A}$. Hence

$$RLF = \varprojlim_I \varinjlim_J F \cong \varprojlim_I (\varinjlim_{j_1/j} F(1\times P)) \cong \varprojlim_{I/i_1} ((\varinjlim_{j_1/J} F(1\times P))Q) \cong \varprojlim_{I/i_1} \varinjlim_{j_1/J} F(Q\times P).$$

Similarly $LRF \cong \varinjlim_{j_1/J} \varprojlim_{I/i_1} F(Q\times P)$. In view of Proposition 1.9, the hypotheses of Corollary 4.15 imply that $F(Q\times P)$ satisfies the hypotheses of Theorem 4.10.

By Theorem 3.6 (dual) $RLF = \mathcal{F}\text{-}\varprojlim_I LF = \mathcal{F}\text{-}\varprojlim_I \varinjlim_J \widetilde{F}$. On the other hand, by the same theorem we have $\mathcal{F}\text{-}\varprojlim_I \widetilde{F}j = RFj$. Let $\mathcal{F}\text{-}\varprojlim_I \widetilde{F}$ be the functor $R\widetilde{F}$. Suppose RF is d-conservative. Then $\varinjlim_J \mathcal{F}\text{-}\varprojlim_I \widetilde{F} = \varinjlim_J \varprojlim_I F = LRF$. Hence we have

Theorem 4.16. *If the functor $F\colon I\times J \to \mathcal{A}$ satisfies the hypotheses of Theorem* 4.10 *and in addition $RF\colon I \to \mathcal{A}$ is* d-*conservative, then*

$$\varinjlim_J \mathcal{F}\text{-}\varprojlim_I \widetilde{F} \cong \mathcal{F}\text{-}\varprojlim_I \varinjlim_J \widetilde{F}.$$

§5. The AB5 case

In Abelian categories satisfying the Grothendieck axiom AB5, direct limits preserve exactness. This fact entails some nice preservation properties for direct limits. In particular we have

Theorem 5.1. *Let $G, H\colon J \to \mathcal{A}$ be functors where J is* qf *and $\mathcal{A}$ is* AB5. *Let $\kappa\colon G \to H$ be a natural transformation. Then if*

$$(5.2)\qquad \begin{tikzcd} Gj \arrow[r, "G\psi"] \arrow[d, "\kappa_j"'] & Gj_2 \arrow[d, "\kappa_{j_2}"] \\ Hj \arrow[r, "H\psi"'] & Hj_2 \end{tikzcd}$$

is exact, cartesian, cocartesian for every $\psi : j \to j_2$, *j-fixed, then also*

$$(5.3)\qquad \begin{tikzcd} Gj \arrow[r, "\lambda_j"] \arrow[d, "\kappa_j"'] & LG \arrow[d, "L\kappa"] \\ Hj \arrow[r, "\pi_j"'] & LH \end{tikzcd}$$

is exact, cartesian, cocartesian, respectively, where $\{LG, \lambda_j\} = \varinjlim G$ *and* $\{LH, \pi_j\} = \varinjlim H$.

Proof. Let the functors $T', T, T'' : j/J \to \mathscr{A}$ be defined by $T' = GjP$, $T = (Hj \oplus G)P$, $T'' = HP$ where Gj, Hj are constant functors $J \to \mathscr{A}$ and $P: j/J \to J$ is the projection functor which we have seen to be cofinal. Recall that with J also j/J is qf. We then have the natural transformations

$$\mu : T' \to T \quad \text{given by} \quad \mu\psi = \{\kappa_j, G\psi\} \quad \text{and}$$

$$\epsilon : T \to T'' \quad \text{given by} \quad \epsilon\psi = \langle H, -\kappa_{G\psi}\rangle$$

which compose into the sequence

$$(5.4)\qquad T' \xrightarrow{\mu} T \xrightarrow{\epsilon} T''.$$

Clearly (5.4) is left exact if and only if (5.2) is exact. From (5.4) we obtain

$$(5.5)\qquad \varinjlim T' \xrightarrow{\varinjlim \mu} \varinjlim T \xrightarrow{\varinjlim \epsilon} \varinjlim T''.$$

Since direct limits in AB5 categories preserve exactness, (5.5) has the same exactness properties as (5.4). More explicitly, (5.5) is

$$\varinjlim GjC \xrightarrow{\{\varinjlim \kappa_j, \varinjlim G\psi\}} \varinjlim HjC \oplus \varinjlim GC \xrightarrow{\langle \varinjlim H\psi, -\varinjlim \kappa_C\rangle} \varinjlim HC$$

and since C is cofinal we see that (5.5) is equal to

$$Gj \xrightarrow{\{\kappa_j, \lambda_j\}} Hj \oplus LG \xrightarrow{\langle \pi_j, -L\kappa\rangle} LH.$$

Clearly (5.5) is left exact if and only if (5.3) is cartesian.

The same argument works if we replace "cartesian" by "cocartesian", "exact", and of course "bicartesian".

When $\mathcal{A}$ is an AB5 category, in view of Theorem 5.1, the hypotheses of Corollary 4.15 become quite unrestrictive. Indeed, we have

Theorem 5.6. *Let I be* qf', *J* qf *and $\mathcal{A}$* AB5. *Let F: $I \times J \to \mathcal{A}$ be a functor for which* (4.2) *is cartesian for every φ: $i \to i_1$, with i_1 fixed and ψ: $j' \to j$ with $\psi = P\psi$ (P: $j_1/J \to J$, j_1 fixed). Then $LRF \cong RLF$*, i.e., $\varinjlim_J \varprojlim_I F \cong \varprojlim_I \varinjlim_J F$.

Proof. Since $\mathcal{A}$ is AB5, F_i is u-conservative for every $i = Q(i \to i_1)$ by Theorem 3.4. Furthermore, by Theorem 5.1 and the remark after Theorem 4.10, (4.4) is cartesian for every φ: $i \to i_1$ and every $j = P(j_1 \to j)$, and (4.11) is cartesian for $i = i_1$ and every $j = P(j_1 \to j)$. Thus F satisfies all the hypotheses of Corollary 4.15, hence $LRF \cong RLF$.

References

[1] H. Bass, Algebraic K-theory (Benjamin, 1968).
[2] B. Eckmann and P.J. Hilton, Commuting limits with colimits, J. Algebra 11 (1969) 116–144.
[3] P.J. Hilton, Correspondences and exact squares, Conf. on Categorical Algebra (La Jolla) 1965 (Springer-Verlag) 254–271.
[4] S. MacLane, Lectures in categorical algebra, Bowdoin College, 1969.
[5] H. Schubert, Kategorien, Heidelberger Taschenbücher (Springer-Verlag, 1970).

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 2 (1971) pp. 199–217.

ON FILTERED SYSTEMS OF GROUPS, COLIMITS, AND KAN EXTENSIONS

Peter J. HILTON
Batelle Research Center, Seattle, Washington
Cornell University, Ithaca, New York

Received February 1971

§1. Introduction

This paper is a sequel to [7]. In that paper we studied direct systems of groups and direct limits. Our particular interest lay in the following question. Suppose given a functor $W_0 : \mathfrak{G}_0 \to \mathfrak{C}$ from a full subcategory $\mathfrak{G}_0$ of the category of groups $\mathfrak{G}$ to a cocomplete category $\mathfrak{C}$, and suppose that G is a group which can be expressed as the direct limit of a direct system of groups in $\mathfrak{G}_0$,

$$G = \varinjlim_{\alpha} G_\alpha, \quad G_\alpha \in \mathfrak{G}_0 .$$

Under what circumstances do we obtain a functor $W_1 : \mathfrak{G}_1 \to \mathfrak{C}$, extending W_0, by setting

$$W_1(G) = \varinjlim_{\alpha} W_0(G_\alpha) ,$$

where $\mathfrak{G}_1$ is the full subcategory of $\mathfrak{G}$ whose objects are precisely such direct limits of groups in $\mathfrak{G}_0$? A second obvious question is to ask when W_1 is the Kan extension of W_0 to $\mathfrak{G}_1$ – this question was referred to briefly in Section 7 of [7], where a sufficient condition, due to Ulmer, was given in case W_0 is additive.

This paper is concerned with this second question, but does not pursue the direction taken by Ulmer. Instead, we exploit the possibility that the category of $\mathfrak{G}_0$-objects over G, G in $\mathfrak{G}_1$, may be a *filtering category* in the sense of Grothendieck-Artin-Mazur [1, 2]; and this category plays a crucial role in the definition of the Kan extension (see Section 6). Thus, we must generalize the theory developed in [7] by replacing directed sets by filtering categories, and much of the technical detail of this paper is devoted to this generalization. Some complication of definitions and demonstrations results, but the basic facts remain valid in this broader context. In particular,

we observe an interesting formal analogy between the theory of $\mathfrak{F}$, the category of filtering categories, and $\mathfrak{C}^{\Sigma}$, the category of filtered systems over $\mathfrak{C}$, on the one hand, and the theory of fibrations in topology, on the other. In fact, we introduce the notion of a *fibre-map* in $\mathfrak{F}$ which we prove to have the crucial properties: (a) the pull-back of a fibre-map is a fibre-map, and (b) every morphism of $\mathfrak{F}$ is canonically expressible as the composite of a 'weak equivalence' and a fibre-map. Of course, fibre-maps had already been introduced into category theory, in an even more general setting; see, e.g., [5].

The development of the theory of colimits in $\mathfrak{C}^{\Sigma}$, and, in particular, in $\mathfrak{G}^{\Sigma}$, where $\mathfrak{G}$ is the category of groups, enables us to obtain an answer to the first question referred to above, exactly corresponding to the answer in [7], but the broader context then allows us to give sufficient conditions for W_1 to be the Kan extension of W_0 (see Theorem 6.1). These conditions make no apparent appeal to the additivity of W_0. On the other hand, we find a local version of Ulmer's criterion ((7.3) of [7]) appearing, in that we ask that, given any G in $\mathfrak{G}_1$, then for some filtering category I and some functor $F: I \to \mathfrak{G}_0$ with $\varinjlim F = G$, we have an equivalence

$$\varinjlim \operatorname{Hom}(G_0, F) \cong \operatorname{Hom}(G_0, \varinjlim F)$$

for all G_0 in $\mathfrak{G}_0$. This condition above is not, however, claimed to be sufficient; it is only sufficient if we know that W_1 exists.

A very minor role is played in the entire theory by the category of groups, and any other category would serve as well, provided it were cocomplete and satisfied the condition that colimits commute with pull-backs (see Theorem 3.4). Thus, the theory should lend itself to substantial generalization in this direction. We give one application, to coefficients in generalized cohomology theories [6]: this application could also have been tackled successfully using Ulmer's criterion.

The author benefited from conversations with Beno Eckmann in developing this approach to the problem. He also learned from Barry Mitchell of another possible approach exploiting the fact that, given a filtering category I, there is a cofinal functor from I to a directed set. This last fact demonstrates, of course, that the category $\mathfrak{G}_1$ over which W_0 is to be extended remains the same whether we confine attention to directed sets or permit arbitrary filtering categories.

This paper constitutes an expanded version of the first part of the author's invited talk at the Nice Congress of Mathematicians [8].

§2. The category of filtered systems and the colimit functor

We recall from [1, 2] the concepts of *filtering category* and *cofinal functor.* Proposition 2.6 of [2] assures us that given cofinal functors

$$I \xrightarrow{T} J \xrightarrow{S} K$$

of filtering categories, ST is again cofinal, so that we may form the category $\mathfrak{F}$ of filtering categories and cofinal functors.

Now let $\mathfrak{C}$ be an arbitrary category. We form the *associated filtered category* or *category of filtered systems*, $\mathfrak{C}^\Sigma$, as follows. An object of $\mathfrak{C}^\Sigma$ is a pair (I, F), often abbreviated to F, where I is a filtering category and $F\colon I \to \mathfrak{C}$ is a functor. A *morphism* $\Phi\colon (I, F) \to (J, G)$ of $\mathfrak{C}^\Sigma$ is a pair $\Phi = (T, u)$, where $T\colon I \to J$ is a cofinal functor and $u\colon F \to GT$ is a natural transformation. *Composition* of morphisms in $\mathfrak{C}^\Sigma$ is defined by

$$(2.1) \qquad (S, v)(T, u) = (ST, vT \circ u)\,.$$

It is easy to check that $\mathfrak{C}^\Sigma$ is a category. If I is any filtering category and $\mathfrak{C}^I$ is the category of objects of $\mathfrak{C}$ filtered by I, that is, the functor category $[I, \mathfrak{C}\,]$, then there is an obvious embedding of $\mathfrak{C}^I$ in $\mathfrak{C}^\Sigma$. In particular, by identifying $\mathfrak{C}$ with $\mathfrak{C}^1$, where 1 is the category with one object and one morphism, we obtain a full embedding

$$(2.2) \qquad P\colon \mathfrak{C} \subseteq \mathfrak{C}^\Sigma\,.$$

Let us henceforth suppose that $\mathfrak{C}$ *is cocomplete*. For each filtering category I we have the full embedding

$$(2.3) \qquad P_I\colon \mathfrak{C} \subseteq \mathfrak{C}^I$$

and P_I has a left-adjoint, left-inverse

$$(2.4) \qquad L_I\colon \mathfrak{C}^I \to \mathfrak{C}$$

with co-unit $\pi_I\colon 1 \to P_I L_I$. If $T\colon I \to J$ is cofinal, define

$$(2.5) \qquad \mathfrak{C}^T\colon \mathfrak{C}^J \to \mathfrak{C}^I$$

by $\mathfrak{C}^T(G) = GT$. Obviously,

$$(2.6) \qquad \mathfrak{C}^T P_J = P_I\,.$$

Theorem 2.10 of [2] asserts that, given L_I, π_I, we may take $L_J^{(T)} = L_I\mathfrak{C}^T$ (so that $L_J^{(T)}P_J = 1$) and we may determine a unique natural transformation $\pi_J^{(T)}\colon 1 \to P_J L_J^{(T)}$ by the equation

$$(2.7) \qquad \mathfrak{C}^T \pi_J^{(T)} = \pi_I \mathfrak{C}^T\,;$$

and that $L_J^{(T)}$ is then left-adjoint, left-inverse to P_J with co-unit $\pi_J^{(T)}$. Suppose now

that, to each filtering category I, we have selected a fixed left-adjoint, left-inverse L_I to P_I with co-unit π_I. Then the uniqueness of left-adjoints implies that, to each cofinal functor T: $I \to J$ in F there is associated a unique natural equivalence

$$(2.8) \qquad \tau_T\colon L_I\mathfrak{C}^T \to L_J \qquad (L_I\mathfrak{C}^T = L_J^{(T)})$$

such that $\tau_T P_J = 1$ and the diagram

$$(2.9) \qquad \begin{tikzcd} 1 \arrow[r, "\pi_J^{(T)}"] \arrow[dr, "\pi_J"'] & P_J L_I \mathfrak{C}^T \arrow[d, "P_J\tau_T"] \\ & P_J L_J \end{tikzcd}$$

commutes, where $\pi_J^{(T)}$ is determined by (2.7). Of course, τ_1 is the identity.

We define L: $\mathfrak{C}^\Sigma \to \mathfrak{C}$ by

$$(2.10) \qquad L(I,F) = L_I(F)\,, \qquad L(T,u) = \tau_T(G) \circ L_I(u)$$

and we define π: $1 \to PL$: $\mathfrak{C}^\Sigma \to \mathfrak{C}^\Sigma$ by

$$(2.11) \qquad \pi(I,F) = (C_I, \pi_I(F))\,,$$

where C_I: $I \to 1$ is the constant functor. The main result of this section is then the following.

Theorem 2.1. *L: $\mathfrak{C}^\Sigma \to \mathfrak{C}$, given by* (2.10), *is a functor, and π: $1 \to PL$, given by* (2.11), *is a natural transformation. Moreover, L is left-adjoint, left-inverse to P with co-unit π.*

Proof. We first remark that, given

$$(2.12) \qquad I \xrightarrow{T} J \xrightarrow{S} K$$

in $\tilde{\mathfrak{F}}$, then

$$(2.13) \qquad \mathfrak{C}^S \pi_K^{(ST)} = \pi_J^{(T)} \mathfrak{C}^S\,.$$

For this equation simply asserts that if we use the process described in Theorem 2.10 of [2] to extend a given choice of the left-adjoint, left-inverse to P_I, through T, to a left-adjoint, left-inverse to P_J, and then extend again, through S, to a left-adjoint, left-inverse to P_K, we obtain the extension of the given choice through ST. From

(2.13) we infer the key identity, again based on (2.12),

$$\tau_{ST} = \tau_S \circ \tau_T \mathfrak{C}^S . \tag{2.14}$$

Since $\tau_S \circ \tau_T \mathfrak{C}^S$ is a natural equivalence $L_I \mathfrak{C}^{ST} = L_I \mathfrak{C}^T \mathfrak{C}^S \to L_K$ and

$$(\tau_S \circ \tau_T \mathfrak{C}^S) P_K = \tau_S P_K \circ \tau_T P_J = 1 ,$$

we have only to prove the commutativity of the diagram

$$\begin{tikzcd} 1 \arrow[r, "\pi_K^{(ST)}"] \arrow[dr, "\pi_K"'] & P_K L_I \mathfrak{C}^{ST} \arrow[d, "P_K(\tau_S \circ \tau_T \mathfrak{C}^S)"] \\ & P_K L_K \end{tikzcd}$$

or, equivalently, the relation

$$\pi_K^{(S)} = P_K \tau_T \mathfrak{C}^S \circ \pi_K^{(ST)} . \tag{2.15}$$

In view of the uniqueness property of (2.7), applied to the relation

$$\mathfrak{C}^S \pi_K^{(S)} = \pi_J \mathfrak{C}^S ,$$

it is sufficient to prove that

$$\pi_J \mathfrak{C}^S = \mathfrak{C}^S P_K \tau_T \mathfrak{C}^S \circ \mathfrak{C}^S \pi_K^{(ST)} ; \tag{2.16}$$

but

$$\begin{aligned} \mathfrak{C}^S P_K \tau_T \mathfrak{C}^S \circ \mathfrak{C}^S \pi_K^{(ST)} &= P_J \tau_T \mathfrak{C}^S \circ \pi_J^{(T)} \mathfrak{C}^S , && \text{by (2.13)}, \\ &= \pi_J \mathfrak{C}^S, && \text{by (2.9)} . \end{aligned}$$

Thus, (2.16) is established and hence (2.15) and (2.14).

We may now readily complete the proof of the theorem. To show that L is a functor, we observe that

$$\begin{aligned} L(S, v)(T, u) = L(ST, vT \circ u) &= \tau_{ST}(H) \circ L_I(vT \circ u) \\ &= \tau_S(H) \circ \tau_T(HS) \circ L_I(vT) \circ L_I(u) , \end{aligned}$$

by (2.14); while

$$L(S,v)\,L(T,u) = \tau_S(H) \circ L_J(v) \circ \tau_T(G) \circ L_I(u)\ ;$$

and $\tau_T(HS) \circ L_I \mathfrak{C}^T(v) = L_J(v) \circ \tau_T(G)$ since τ_T is a natural transformation $L_I \mathfrak{C}^T \to L_J$ and $v\colon G \to HS$, so that v is a morphism of $\mathfrak{C}^J$. Thus, L is a functor and it is plain that $LP = 1$.

We turn attention to π, defined by (2.11). Write $PX = (1, X_1)$, X in $\mathfrak{C}$, where X_1 is the functor taking the value X at the object of 1; similarly write $Pf = (1, f_1)$. Then $PL(I,F) = (1, L_I(F)_1)$ and

$$(C_I, \pi_I(F))\colon (I,F) \to (1, L_I(F)_1)$$

since $L_I(F)_1 C_I = P_I L_I(F)$. Moreover, it is plain that $L\pi = 1$, $\pi P = 1$, in view of the corresponding properties of π_I. Finally we have only to check the naturality of π, that is, the commutativity of the diagram

$$\begin{array}{ccc}
(I,F) & \xrightarrow{(C_I,\,\pi_I(F))} & (1, L_I(F)_1) \\
\Big\downarrow{\scriptstyle (T,u)} & & \Big\downarrow{\scriptstyle (1,f_1)} \\
(J,G) & \xrightarrow[(C_J,\,\pi_J(G))]{} & (1, L_J(G)_1)
\end{array}\quad ,$$

where $f = \tau_T(G) \circ L_I(u)$. Now

$$(1,f_1) \circ (C_I, \pi_I(F)) = (C_I, f_1 C_I \circ \pi_I(F)) = (C_I, P_I f \circ \pi_I(F))$$

$$= (C_I, P_I \tau_T(G) \circ P_I L_I(u) \circ \pi_I(F))$$

while

$$(C_J, \pi_J(G)) \circ (T,u) = (C_J T, \pi_J(G) T \circ u) = (C_I, \mathfrak{C}^T \pi_J(G) \circ u)\,.$$

But

$$\begin{aligned}
\mathfrak{C}^T \pi_J(G) \circ u &= \mathfrak{C}^T P_J \tau_T(G) \circ \mathfrak{C}^T \pi_J^{(T)}(G) \circ u\,, && \text{by (2.9),}\\
&= P_I \tau_T(G) \circ \pi_I \mathfrak{C}^T(G) \circ u\,, && \text{by (2.6) and (2.7)},\\
&= P_I \tau_T(G) \circ \pi_I(GT) \circ u\,,\\
&= P_I \tau_T(G) \circ P_I L_I(u) \circ \pi_I(F)\,,
\end{aligned}$$

since π_I: $1 \to P_I L_I$: $\mathfrak{C}^I \to \mathfrak{C}^I$ is natural and u: $F \to GT$ is a morphism of $\mathfrak{C}^I$. This completes the proof of the theorem.

Remark 1. The conclusion of Theorem 2.1 amounts to a *coherence* statement about colimits over variable index categories. It permits us, in the arguments used in this paper, to assume for convenience that if T: $I \to J$ is a morphism of $\mathfrak{F}$, then $L_J = L_I \mathfrak{C}^T$ and $\pi_J = \pi_J^{(T)}$. This will avoid unnecessarily complicated enunciations.

Remark 2. Generalizations are, of course, possible. We could dispense with the requirement that the index categories $I, J, \ldots$, be filtering, or with the requirement that the functors T be cofinal – or even, perhaps, both. The generality we achieve from considering $\mathfrak{C}^{\Sigma}$ is adequate to our purposes in this paper, but it may well be sensible to relax the conditions to obtain a more comprehensive theory (for example, to express cocompleteness by means of the existence of a single left-adjoint).

§3. Fibre-maps and pull-backs

Definition 3.1. We say that T: $I \to J$ in $\mathfrak{F}$ is a *fibre-map* if, given i in I and ψ: $T(i) \to j'$ in J, there exists ϕ: $i \to i'$ in I with $T(\phi) = \psi$. We say that (T, u): $(I, F) \to (J, G)$ in $\mathfrak{C}^{\Sigma}$ is a *fibre-map* if T is a fibre-map. Notice that (i) a fibre-map T: $I \to J$ is automatically cofinal, and (ii) the fibre-maps form a subcategory of $\mathfrak{F}$.

Now let

$$\begin{tikzcd} & I \arrow[d, "T"] \\ J \arrow[r, "S"'] & K \end{tikzcd}$$

be a diagram in $\mathfrak{F}$. By taking pull-backs in the category of sets, we obtain a commutative diagram of small categories and functors

(3.1)
$$\begin{tikzcd} A \arrow[r, "S'"] \arrow[d, "T'"] & I \arrow[d, "T"] \\ J \arrow[r, "S"'] & K \end{tikzcd} .$$

We prove:

Theorem 3.1. *If T in* (3.1) *is a fibre-map in* $\mathfrak{F}$, *then*

(a) *A is filtering;*

(b) *T' is a fibre-map in* $\mathfrak{F}$;

(c) *S' is in* $\mathfrak{F}$;

(d) (3.1) *is a pull-back in* $\mathfrak{F}$.

Proof (a). We make essential use here and later of Corollary 2.9 of [2], which we quote here in the following form.

Lemma 3.2. *If $S: I \to K$ is in $\mathfrak{F}$ and if $\theta: S(i) \to k$ is in K, there exist $\phi: i \to i'$ in I, $\theta': k \to S(i')$ in K with $\theta'\theta = S(\phi)$.*

Returning to (a), note that $(i, j) \in A$ iff $T(i) = S(j)$ and (ϕ, ψ) is a morphism of A iff $T(\phi) = S(\psi)$. Now suppose that $(i_1, j_1), (i_2, j_2) \in A$. Since I is filtering, we find

$$\begin{tikzcd} i_1 \arrow[dr, "\phi_1"] & \\ & i \\ i_2 \arrow[ur, "\phi_2"'] & \end{tikzcd}$$

in I. Consider

$$\begin{tikzcd} S(j_1) \arrow[dr, "T\phi_1"] & \\ & T(i) \\ S(j_2) \arrow[ur, "T\phi_2"'] & \end{tikzcd}$$

in K. By Lemma 3.2 we find $\theta_q: T(i) \to S(j'_q)$, $\psi_q: j_q \to j'_q$, with $\theta_q T(\phi_q) = S(\psi_q)$, $q = 1, 2$. Since S is cofinal we find $\psi'_q: j'_q \to j'$, $q = 1, 2$, with $S(\psi'_1)\theta_1 = S(\psi'_2)\theta_2$. Set $\theta = S(\psi'_1)\theta_1$. Then, since T is a fibre-map, we find $\phi: i \to i'$ with $T(\phi) = \theta$. We claim that

$$(\phi\phi_q, \psi_q\psi'_q): (i_q, j_q) \to (i', j')$$

is in A for $q = 1, 2$. For

$$T(\phi\phi_q) = T(\phi)T(\phi_q) = S(\psi'_q)\theta_q T(\phi_q) = S(\psi'_q)S(\psi_q) = S(\psi'_q\psi_q).$$

Now let

$$(i_1, j_1) \underset{(\phi_2, \psi_2)}{\overset{(\phi_1, \psi_1)}{\rightrightarrows}} (i_2, j_2)$$

be in A, so that $T(\phi_q) = S(\psi_q)$, $q = 1, 2$. Equalize ϕ_1 and ϕ_2 by $\phi: i_2 \to i_0$, $\phi\phi_1 = \phi\phi_2$, and consider

$$S(j_1) \underset{S\psi_2}{\overset{S\psi_1}{\rightrightarrows}} S(j_2) \xrightarrow{T\phi} T(i_0)$$

in K. By Lemma 3.2 we find $\psi: j_2 \to j$ in J, $\theta: T(i_0) \to S(j)$ in K with $\theta T(\phi) = S(\psi)$.

Equalize $\psi\psi_1$ and $\psi\psi_2$ with $\psi'\colon j \to j'$, $\psi'\psi\psi_1 = \psi'\psi\psi_2$. Since T is a fibre-map, we may find $\phi'\colon i_0 \to i'$ with $T(\phi') = S(\psi')\theta$. Then, plainly, $(\phi'\phi, \psi'\psi)$ is in A and equalizes (ϕ_1, ψ_1) and (ϕ_2, ψ_2). This proves (a).

Proof (b). We suppose given $(i,j) \in A$ and $\psi\colon j \to j'$ in J. Then $S(\psi)\colon T(i) \to S(j')$ in K so that, T being a fibre-map, we have $\phi\colon i \to i'$ in I with $T(\phi) = S(\psi)$. Thus $(\phi, \psi)\colon (i,j) \to (i', j')$ in A with $T'(\phi, \psi) = \psi$.

Proof (c). Given $i \in I$, we find $\theta\colon T(i) \to S(j')$ in K since S is cofinal and $\phi\colon i \to i'$ in I with $T(\phi) = \theta$ since T is a fibre-map. Thus $(i', j') \in A$ and $\phi\colon i \to S'(i', j')$.

Now suppose given $(i_2, j_2) \in A$ and $i_1 \overset{\phi_1}{\underset{\phi_2}{\rightrightarrows}} i_2$ in I. We argue just as for the second part of (a), as far as the construction of ϕ, ψ, θ with $\phi\phi_1 = \phi\phi_2$, $\theta T(\phi) = S(\psi)$. Now construct $\bar{\phi}\colon i_0 \to i$ in I with $T(\bar{\phi}) = \theta$. Then $(\bar{\phi}\phi, \psi)$ is in A and $S'(\bar{\phi}\phi, \psi) = \bar{\phi}\phi$ equalizes ϕ_1 and ϕ_2.

Proof (d). Let $U\colon L \to I$, $V\colon L \to J$ be in $\mathfrak{F}$ with $TU = SV$. There is plainly a unique functor $W\colon L \to A$ with $S'W = U$, $T'W = V$, namely, $W(\ell) = (U\ell, V\ell)$, $W(\sigma) = (U\sigma, V\sigma)$. It remains to show that W is cofinal. Let $(i,j) \in A$. Since U is cofinal we find $\phi\colon i \to U\ell$ in I. Consider $T\phi\colon Sj \to SV\ell$. Since S is cofinal we may find, by Proposition 2.8 of [2], $\psi\colon j \to j'$, $\psi_1\colon V\ell \to j'$ in J with $S\psi_1 \circ T\phi = S\psi$. Since V is cofinal we may find, by Lemma 3.2, $\sigma\colon \ell \to \ell'$ in L, $\psi'\colon j' \to V\ell'$ in J with $\psi'\psi_1 = V\sigma$. Set $\phi_0 = U\sigma \circ \phi\colon i \to U\ell'$, $\psi_0 = \psi'\psi\colon j \to V\ell'$. Then $T\phi_0 = TU\sigma \circ T\phi = SV\sigma \circ T\phi = S\psi' \circ S\psi_1 \circ T\phi = S\psi' \circ S\psi = S\psi_0$. Thus (ϕ_0, ψ_0) is in A and $(\phi_0, \psi_0)\colon (i,j) \to (U\ell', V\ell')$.

Now let $(i,j) \overset{(\phi_1, \psi_1)}{\underset{(\phi_2, \psi_2)}{\rightrightarrows}} (U\ell, V\ell)$ be given in A. Since U is cofinal we find $\sigma'\colon \ell \to \ell'$ in L with $U\sigma' \circ \phi_1 = U\sigma' \circ \phi_2$. Since V is cofinal we find $\sigma''\colon \ell \to \ell''$ in L with $V\sigma'' \circ \psi_1 = V\sigma'' \circ \psi_2$. Since L is filtering we find $\tau'\colon \ell' \to \bar{\ell}$, $\tau''\colon \ell'' \to \bar{\ell}$ with $\tau'\sigma' = \tau''\sigma'' = \chi\colon \ell \to \bar{\ell}$. Then $(U\chi, V\chi)$ equalizes (ϕ_1, ψ_1) and (ϕ_2, ψ_2) so that W is cofinal.

This completes the proof of the theorem.

Now let $\mathfrak{C}$ admit pull-backs, and let
$$\begin{array}{ccc} & & (I,F) \\ & & \big\downarrow{\scriptstyle (T,u)} \\ (J,G) & \xrightarrow[(S,v)]{} & (K,H) \end{array}$$
be a diagram in $\mathfrak{C}^{\Sigma}$ with (T, u) is a fibre-map. Using (3.1) we construct

$$(3.2)\qquad \begin{array}{ccc} (A,E) & \xrightarrow{(S',v')} & (I,F) \\ {\scriptstyle (T',u')}\big\downarrow & & \big\downarrow{\scriptstyle (T,u)} \\ (J,G) & \xrightarrow[(S,v)]{} & (K,H) \end{array}$$

where, for $(i,j) \in A$, $E(i,j)$ is the pull-back, in $\mathfrak{C}$, of the diagram

$$\begin{tikzcd} E(i,j) \arrow[r, dashed, "v'(i,j)"] \arrow[d, dashed, "u'(i,j)"'] & F(i) \arrow[d, "u(i)"] \\ G(j) \arrow[r, "v(j)"] & HT(i) = HS(j), \end{tikzcd}$$

with the obvious value of E on the morphisms of A.

Theorem 3.3. *If (T, u) is a fibre-map, then* (3.2) *is a pull-back in* $\mathfrak{C}^{\Sigma}$.
The proof of this theorem is entirely straightforward and will be omitted.

We now prove the only theorem which involves a further restriction on the category $\mathfrak{C}$. We will, in fact, prove the theorem for $\mathfrak{C} = \mathfrak{G}$, the category of groups; we will then discuss the more general case in remarks following the proof.

Theorem 3.4. *The colimit in $\mathfrak{G}^{\Sigma}$ commutes with pull-backs of fibre-maps.*

Proof. The enunciation requires us to prove the following. Consider the pull-back diagram (3.2) in $\mathfrak{G}^{\Sigma}$ and form the diagram, in $\mathfrak{G}$,

(3.3)
$$\begin{tikzcd} & L(I,F) \arrow[d, "L(T,u)"] \\ L(J,G) \arrow[r, "L(S,v)"] & L(K,H) . \end{tikzcd}$$

Now take the pull-back

(3.4)
$$\begin{tikzcd} Q \arrow[r, "\sigma"] \arrow[d, "\tau"'] & L(I,F) \arrow[d, "L(T,u)"] \\ L(J,G) \arrow[r, "L(S,v)"] & L(K,H) \end{tikzcd}$$

of (3.3). Then we must show that there is an equivalence $\omega: L(A,E) \cong Q$ such that $\tau\omega = L(T', u')$, $\sigma\omega = L(S', v')$, provided that T is a fibre-map. We prove this by recalling the explicit construction of colimits in $\mathfrak{G}^{\Sigma}$ (see, e.g., [3]) and the familiar criterion for a colimit in the category of groups (see [7]).

Let $(I, F) \in {}^{\Sigma}$. In the set $\bigcup_{i \in I} F(i)$ set up an equivalence relation $x_{i_1} \sim x_{i_2}$ to mean that there exists

$$i_1 \xrightarrow{\phi_1} i \xleftarrow{\phi_2} i_2$$

in I with $\phi_1(x_{i_1}) = \phi_2(x_{i_2})$. Let $[x_i]$ be the equivalence class of x_i. Multiply the equivalence classes by the rule

$$(3.5) \qquad [x_i][x_{i'}] = [\phi(x_i)\,\phi'(x_{i'})]$$

where $i \xrightarrow{\phi} i_0 \xleftarrow{\phi'} i'$ is in I. Then the equivalence classes, together with (3.5), yield a group which may be identified with $L(I, F) = L_I(F)$. Moreover, $\pi_i\colon F(i) \to L(I, F)$ is then given by $\pi_i(x_i) = [x_i]$. If $(T, u)\colon (I, F) \to (J, G)$ in $\mathfrak{G}^{\Sigma}$, then $L(T, u)\colon L(I, F) \to$
$\to L(J, G)$ is given by

$$(3.6) \qquad L(T, u)[x_i] = [u_i(x_i)]\,, u_i(x_i) \in GT(i)\,.$$

Now suppose given $\kappa_i\colon F(i) \to M$ in $\mathfrak{G}$ with $\kappa_{i_2}F(\phi) = \kappa_{i_1}$ for all $\phi\colon i_1 \to i_2$ in I. Then the system $(M; \kappa_i)$ is equivalent to $(L(I, F); \pi_i)$ if and only if it has the following two properties:

(a) each $m \in M$ belongs to some $\kappa_i F(i)$;

(b) if $\kappa_i(x) = e$, $x \in F(i)$, then there exists $\phi\colon i \to i'$ such that $F(\phi)x = e$.

We use this criterion to prove the theorem. There is an evident map

$$\kappa_{ij}\colon E(i,j) \to Q\,, \quad i \in I,\ j \in J,\ T(i) = S(j)\,,$$

given by

$$(3.7) \qquad \kappa_{ij}(x, y) = ([x], [y]), \quad x \in F(i),\ y \in G(j),\ u_i(x) = v_j(y)\,,$$

and, plainly, $\kappa_{i'j'}E(\phi, \psi) = \kappa_{ij}$ for $\phi\colon i \to i'$, $\psi\colon j \to j'$, $T(\phi) = S(\psi)$. Thus it only remains to establish that conditions (a), (b) above hold for the maps κ_{ij}. Let $([x], [y]) \in Q$, $x \in F(i)$, $y \in G(j)$, $[u_i(x)] = [v_j(y)]$. Since S is cofinal and J is filtering, we can plainly find $\psi\colon j \to j'$ in J, $\theta\colon T(i) \to S(j')$ in K; and since T is a fibre-map we may find $\phi\colon i \to i'$ in I with $T(\phi) = \theta$. Then if $x' = F(\phi)x$, $y' = G(\psi)y$, we have $[x'] = [x]$, $[y'] = [y]$; thus we may assume without real loss of generality that $x \in F(i)$, $y \in G(j)$ *and* $T(i) = S(j)$. Since $[u_i(x)] = [v_j(y)]$, there exists $\theta\colon T(i) = S(j) \to k$ in K with $H(\theta)u_i(x) = H(\theta)v_j(y)$. By Lemma 3.2 there exists $\theta'\colon k \to S(j')$ in K with $\theta'\theta = S(\psi)$ for some $\psi\colon j \to j'$ in J. Since T is a fibre-map, there exists $\phi\colon i \to i'$ in I with $T(\phi) = S(\psi)$. Then if $x' = F(\phi)x$, $y' = G(\psi)y$, we have $[x'] = [x]$, $[y'] = [y]$ and

$$u_{i'}(x') = u_{i'}F(\phi)x = HT(\phi)u_i(x) = H(\theta')\,H(\theta)\,u_i(x)\,,$$

$$v_{j'}(y') = v_{j'}G(\psi)y = HS(\psi)v_j(y) = H(\theta')H(\theta)v_j(y)\,,$$

so that $u_{i'}(x') = v_{j'}(y')$. Thus $(x', y') \in E(i', j')$ and $\kappa_{i'j'}(x', y') = ([x'], [y']) =$
$= ([x], [y])$. This proves (a). To prove (b), suppose $\kappa_{ij}(x, y) = e$, so that $[x] = e$, $[y] = e$, $x \in F(i)$, $y \in G(j)$, $T(i) = S(j)$. Then there exist $\phi\colon i \to i'$ in I, $\psi\colon j \to j'$ in J

with $F(\phi)x = e$, $G(\psi)y = e$. Since K is filtering we may find $\theta_1 \colon Sj' \to k$, $\theta_2 \colon Ti' \to k$ in K with $\theta_1 S\psi = \theta_2 T\phi$. By Lemma 3.2, we may find $\psi' \colon j' \to j''$, $\theta \colon k \to Sj''$ with $\theta\theta_1 = S\psi'$. Finally, since T is a fibre-map, we find $\phi' \colon i' \to i''$ with $T\phi' = \theta\theta_2$. Then

$$T(\phi'\phi) = \theta\theta_2 T\phi = \theta\theta_1 S\psi = S(\psi'\psi),$$

and $F(\phi'\phi)x = e$, $G(\psi'\psi)y = e$. Thus $E(\phi'\phi, \psi'\psi)(x, y) = e$ so that (b) is proved and, with it, Theorem 3.4.

Remarks. In our subsequent arguments we may replace the category $\mathfrak{G}$ by any category $\mathfrak{C}$ such that colimits in $\mathfrak{C}^{\Sigma}$ commute with pull-backs of fibre-maps. It is not difficult to show that $\mathfrak{C}$ has this property if and only if, for all filtering categories I, colimits in $\mathfrak{C}^{I}$ commute with pull-backs. Among such categories are to be found the category of sets and the category of Λ-modules for Λ a unitary ring.

§4. The canonical factorization

In this section we introduce the canonical factorization of a morphism of $\mathfrak{C}^{\Sigma}$. Thus we suppose given

$$(T, u) \colon (I, F) \to (J, G)$$

in $\mathfrak{C}^{\Sigma}$ and we proceed to define the objects and the morphisms of the diagram

$$(4.1) \qquad \begin{tikzcd} & (\bar{I}, \bar{F}) \arrow[d, "(\bar{T}, \bar{u})"] \arrow[dl, shift left, "(R', v')"] \\ (I, F) \arrow[ur, shift left, "(R, v)"] \arrow[r, "(T, u)"'] & (J, G) \end{tikzcd} .$$

First, $\bar{I}$ is the category whose objects are pairs (i, ψ), where $\psi \colon T(i) \to j$ in J, and whose morphisms $(i, \psi) \to (i', \psi')$ are pairs $\phi \colon i \to i'$ in I, $\theta \colon j \to j'$ in J such that the diagram

$$\begin{tikzcd} T(i) \arrow[r, "\psi"] \arrow[d, "T\phi"'] & j \arrow[d, "\theta"] \\ T(i') \arrow[r, "\psi'"] & j' \end{tikzcd}$$

commutes. Composition of morphisms in $\bar{I}$ is just component-wise, and $\bar{I}$ is plainly a small category. We define $\bar{T}$: $\bar{I} \to J$ by

$$(4.2) \qquad \bar{T}(i, \psi) = j, \qquad \bar{T}(\phi, \theta) = \theta\,.$$

Proposition 4.1. *$\bar{I}$ is a filtering category and $\bar{T}$ is a fibre-map.*

Proof. Suppose given (i_q, ψ_q), ψ_q: $T(i_q) \to j_q$, $q = 1, 2$. We first find $i_1 \xrightarrow{\phi_1} i \xleftarrow{\phi_2} i_2$ in I. Since J is filtering, we find commutative squares

$$\begin{tikzcd} T(i_q) \arrow[r, "\psi_q"] \arrow[d, "T\phi_q"] & j_q \arrow[d, "\theta_q"] \\ T(i) \arrow[r, "\psi'_q"'] & j'_q \end{tikzcd}$$

$$\begin{tikzcd} T(i) \arrow[r, "\psi'_1"] \arrow[d, "\psi'_2"] \arrow[dr, dashed, "\psi"] & j'_1 \arrow[d, "\theta'_1"] \\ j'_2 \arrow[r, "\theta'_2"'] & j \end{tikzcd}$$

in J and we have $(\phi_q, \theta'_q\theta_q)$: $(i_q, \psi_q) \to (i, \psi)$, $q = 1, 2$.

Now consider the commutative squares

$$\begin{tikzcd} T(i) \arrow[r, "\psi"] \arrow[d, shift right, "T\phi_1"'] \arrow[d, shift left, "T\phi_2"] & j \arrow[d, shift right, "\theta_1"'] \arrow[d, shift left, "\theta_2"] \\ T(i') \arrow[r, "\psi'"] & j' \end{tikzcd}$$

Equalize ϕ_1, ϕ_2 with ϕ: $i' \to i_1$, $\phi\phi_1 = \phi\phi_2$, and equalize θ_1, θ_2 with θ': $j' \to j_0$, $\theta'\theta_1 = \theta'\theta_2$. Since J is filtering, we may construct θ'': $j_0 \to j_1$, ψ_1: $T(i_1) \to j_1$ with $\psi_1 T\phi = \theta''\theta'\psi'$. Then if $\theta = \theta''\theta'$, (ϕ, θ): $(i', \psi') \to (i_1, \psi_1)$ and (ϕ, θ) equalizes (ϕ_1, θ_1), (ϕ_2, θ_2). This shows that $\bar{I}$ is filtering.

We now show that $\bar{T}$ is a fibre-map. We are given (i, ψ) in $\bar{I}$ and θ: $j \to j'$ in J. Then $(1, \theta)$: $(i, \psi) \to (i, \theta\psi)$ in $\bar{I}$ and $\bar{T}(1, \theta) = \theta$. This proves the proposition.

We now continue to construct (4.1). We set $R(i) = (i, 1_{Ti})$, $R(\phi) = (\phi, T\phi)$, $R'(i, \psi) = i$, $R'(\phi, \theta) = \phi$, and prove

Proposition 4.2. *R and R' are cofinal. Moreover $R'R = 1$ and $\bar{T}R = T$.*

Proof. We prove R cofinal; the remaining assertions are trivial. Given (i, ψ) in $\bar{I}$, we find $\theta: j \to T(i'')$ since T is cofinal. By Proposition 2.8 of [2] we now find $\phi': i'' \to i'$, $\phi: i \to i'$ such that $T(\phi')\theta\psi = T(\phi)$. Then $(\phi, T(\phi')\theta): (i, \psi) \to (i', 1_{Ti'})$.

Now let $(\phi_q, \theta_q): (i, \psi) \to (i', \psi) \to (i', 1_{Ti'})$, $q = 1, 2$. Since T is cofinal, we may equalize θ_1, θ_2 with $T\phi', \phi': i' \to i''$, $T(\phi')\theta_1 = T(\phi')\theta_2$. Now equalize $\phi'\phi_1, \phi'\phi_2$ with $\phi'': i'' \to i'''$, $\phi''\phi'\phi_1 = \phi''\phi'\phi_2$. If $\phi = \phi''\phi'$, then $\phi: i' \to i'''$ and $R\phi$ equalizes (ϕ_1, θ_1) and (ϕ_2, θ_2).

Note that Proposition 4.2 provides an alternative proof that $\bar{I}$ is filtering.

We continue the description of (4.1) by setting $\bar{F}(i, \psi) = F(i)$, $\bar{F}(\phi, \theta) = F(\phi)$, $\bar{u}(i, \psi) = G\psi \circ u(i)$, $\bar{v} = 1$, $v' = 1$. It is easy to check that $\bar{u}$ is a natural transformation, $\bar{u}: \bar{F} \to G\bar{T}$. Moreover,

$$(4.3) \qquad (R', v')(R, v) = (1, 1), \quad (\bar{T}, \bar{u})(R, v) = (T, u).$$

This completes the description of the canonical factorization. We observe that $L(R, v)$, $L(R', v')$ are mutually inverse equivalences $L(I, F) \cong L(\bar{I}, \bar{F})$; we will feel free in the sequel to identify $L(I, F)$ with $L(\bar{I}, \bar{F})$, thereby identifying $L(T, u)$ with $L(\bar{T}, \bar{u})$.

Remarks. For our purposes the notion of a fibre-map is adequate; however to obtain good 'lifting' theorems, one would need a stronger notion. If $T: I \to J$ is in $\mathfrak{F}$ we may describe a *lift* for T as a function

$$\Theta: \bar{I} \to \text{Morph}\, I,$$

such that (i) $D\Theta(i, \psi) = i$, where $D\phi$ denotes the domain of ϕ; (ii) $T\Theta(i, \psi) = \psi$; (iii) $\Theta R(i) = 1_i$. One may then show that if T has lift Θ and S has lift Φ, then ST has lift Ψ, where

$$\Psi(i, \chi) = \Theta(i, \Phi(T(i), \chi)), \quad \chi: STi \to k \text{ in } K.$$

If we write $\Psi = \Theta \mathbin{\square} \Phi$, then $\square$ is an associative operation. Plainly, a cofinal functor admitting a lift is a fibre-map, and so the collection of such functors yields a subcategory of the category of fibre-maps (plainly, the identity is the unique lift of the identity). Now if we revert to (3.1) and suppose that T has the lift Θ, then T' has the lift Θ', given by

$$\Theta'((i, j), \psi) = (\Theta(i, S\psi), \psi), \qquad T(i) = S(j), \quad \psi: j \to j'.$$

Finally we remark that $\bar{T}$ in (4.1) has a canonical lift, namely,

$$\Theta((i, \psi), \theta) = (1, \theta), \quad \theta: j \to j'.$$

Thus we could have carried out our work thus far for 'liftable' functors instead of fibre-maps. This is, however, presumably, a narrower class. It is possible, however, that to develop a sensible homotopy theory, one should impose further restrictions on the lift Θ, for example, to require its functoriality.

§5. Filtered systems of groups

Let $\mathfrak{G}_0$ be a full subcategory of the category of groups $\mathfrak{G}$, and consider the colimit functor

$$(5.1) \qquad L: \mathfrak{G}_0^{\Sigma} \to \mathfrak{G}\,.$$

Let $\mathfrak{G}_1$ be the full subcategory of $\mathfrak{G}$ whose objects are precisely the images of L in (5.1). Thus $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \subseteq \mathfrak{G}$. We will make the crucial hypothesis that, given

$$(5.2) \qquad \begin{tikzcd} & G_0 \arrow[d] \\ G_0' \arrow[r] & G_1 \end{tikzcd}$$

in $\mathfrak{G}_1$ with G_0, G_0' in $\mathfrak{G}_0$, then the pull-back of (5.2) is also in $\mathfrak{G}_0$. We call this *hypothesis* P. Now pass to the category of fractions with respect to (5.1), now interpreted as

$$L: \mathfrak{G}_0^{\Sigma} \to \mathfrak{G}_1\,.$$

Thus we obtain the canonical factorization of L as

$$(5.3) \qquad \begin{tikzcd} \mathfrak{G}_0^{\Sigma} \arrow[rr, "L"] \arrow[dr, "Q"'] & & \mathfrak{G}_1 \\ & \widetilde{\mathfrak{G}}_0^{\Sigma} \arrow[ur, "\widetilde{L}"'] & \end{tikzcd}$$

where $\widetilde{\mathfrak{G}}_0^{\Sigma}$ is the category of fractions. Our main theorem in this section is

Theorem 5.1. *Given hypothesis* P, *then* $\widetilde{L}$ *is full and faithful.*

Proof. In view of the preparation provided in Sections 3 and 4, the proof is formally the same as that of Theorem 3.5 of [7]. Thus it is not necessary to repeat the argument here. Moreover, we may pass, as in [7], to the immediate corollary:

Corollary 5.2. *Let* $\mathfrak{C}$ *be a cocomplete category and let* $W_0: \mathfrak{G}_0 \to \mathfrak{C}$ *be a functor. Then, granted hypothesis* P, W_0 *extends to a unique functor* $W_1: \mathfrak{G}_1 \to \mathfrak{C}$ *such*

that

$$(5.4)\qquad W_1 L = LW_0^{\Sigma} : \mathfrak{G}_0^{\Sigma} \to \mathfrak{C}$$

if and only if, for any morphism Φ in $\mathfrak{G}_0^{\Sigma}$, $LW_0^{\Sigma}(\Phi)$ is an equivalence if $L(\Phi)$ is an equivalence.

The argument is again just as in [7] (with some small changes of notation!). We recall in particular that, in proving Theorem 3.5 of [7], we exhibited *minimum paths* for representing morphisms of $\mathfrak{G}_1$; in our current context this means that given $(I_1, F_1), (I_2, F_2)$ in $\mathfrak{G}_0^{\Sigma}$ and a morphism $\phi\colon L_{I_1}(F_1) \to L_{I_2}(F_2)$ in $\mathfrak{G}_1$, there is a preferred diagram

$$(5.5)\qquad (I_1, F_1) \xrightarrow{\Phi_1} (I, F) \xleftarrow{\Phi_2} (I_2, F_2)$$

in $\mathfrak{G}_0^{\Sigma}$ such that $L(\Phi_1) = \phi$, $L(\Phi_2) = 1$; and, given any other diagram

$$(I_1, F_1) \xrightarrow{\Phi_1'} (I', F') \xleftarrow{\Phi_2'} (I_2, F_2)$$

in $\mathfrak{G}_0^{\Sigma}$ such that $L(\Phi_1') = \phi$, $L(\Phi_2') = 1$, there is a unique $\Psi\colon (I', F') \to (I, F)$ with $L(\Psi) = 1$, $\Psi\Phi_q' = \Phi_q$, $q = 1, 2$. We also recall that we do *not* have here a calculus of fractions in the sense of [4].

Remark 1. Hypothesis P is clearly satisfied for any Serre class of Abelian groups.

Remark 2. It is perfectly possible, and straightforward, to generalize the content of Section 4 of [7] to our present case of filtered systems of groups. Thus we would get results which would be applicable if hypothesis P were violated. We do not make these results explicit, since our emphasis here is on the application of Corollary 5.2.

§6. Relation to the Kan extension

We again consider $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \subseteq \mathfrak{G}$ as in Section 5, and construct the Kan extension of $W_0\colon \mathfrak{G}_0 \to \mathfrak{C}$ to $\mathfrak{G}$. Given G in $\mathfrak{G}$, the category I_G of $\mathfrak{G}_0$-objects over G has, as objects, homomorphisms $\chi\colon G_0 \to G$, with G_0 in $\mathfrak{G}_0$, and, as morphisms $\phi\colon \chi \to \chi'$, homomorphisms $\phi\colon G_0 \to G_0'$ rendering commutative the diagram

$$\begin{tikzcd} G_0 \arrow[rr, "\phi"] \arrow[dr, "\chi"'] & & G_0' \arrow[dl, "\chi'"] \\ & G & \end{tikzcd}$$

We define $W_G: I_G \to \mathfrak{C}$ by $W_G(\chi) = W_0(G_0)$, $W_G(\phi) = W_0(\phi)$; and denote by U the underlying functor $U: I_G \to \mathfrak{G}_0$, so that

$$(6.1) \qquad W_G = W_0 U .$$

If I_G is filtering and we write L for the colimit functor, $\mathfrak{C}$ being assumed co-complete, then the *Kan extension* W of W_0 is given by

$$W(G) = L(W_G), \quad G \text{ in } \mathfrak{G},$$

and $W(G) = L(W_0 U) = L W_0^{\Sigma}(U)$. Thus, if $\mathfrak{G}_0$ satisfies hypothesis P and G belongs to $\mathfrak{G}$, then, by (5.4), we infer that if W_0 satisfies the (necessary and sufficient) condition of Corollary 5.2, W_1 exists and

$$(6.2) \qquad W(G) = W_1 L(U) .$$

Let $G \in \mathfrak{G}_1$ so that $G = L(I, F)$, for some (I, F) in $\mathfrak{G}_0^{\Sigma}$. There is then a canonical functor

$$(6.3) \qquad T_G: I \to I_G ,$$

given by $T_G(i) = \pi_I(i)$, $T_G(\phi) = F(\phi)$. We may now prove our main theorem; the first part of the enunciation merely reproduces the essence of Corollary 5.2.

Theorem 6.1. *Consider $\mathfrak{G}_0 \subseteq \mathfrak{G}_1 \subseteq \mathfrak{G}$ as in Section 5 with $\mathfrak{G}_0$ satisfying hypothesis P and assume that $W_0: \mathfrak{G}_0 \to \mathfrak{C}$ has the property that, for any morphism Φ in $\mathfrak{G}_0$, $L W_0^{\Sigma}(\Phi)$ is an equivalence if $L(\Phi)$ is an equivalence. Then W_0 extends to a unique functor $W_1: \mathfrak{G}_1 \to \mathfrak{C}$ such that $W_1 L = L W_0^{\Sigma}: \mathfrak{G}_0^{\Sigma} \to \mathfrak{C}$. Let $G = L(I, F)$ and suppose that $T_G: I \to I_G$ is cofinal. Then $W_1(G) = W(G)$, where W is the Kan extension of W_0 to $\mathfrak{G}_1$.*

Proof. We first note that by [1] or [2; Prop. 2.4], I_G is filtering. Second, it is easy to check that $UT_G = F$, so that $L(U) = L(UT_G) = G$, since T_G is cofinal. Thus, we merely have to apply (6.2) to obtain the result.

Let us interpret the condition that T_G is cofinal. Given (I, F) in $\mathfrak{G}_0^{\Sigma}$ with $L(I, F) = G$, and G_0 in $\mathfrak{G}_0$, we have an evident induced functor $\operatorname{Hom}(G_0, F): I \to \mathfrak{S}$, the category of sets. The morphisms $\pi(i): F(i) \to G$ plainly induce morphisms $\operatorname{Hom}(G_0, F)(i) = \operatorname{Hom}(G_0, F(i)) \to \operatorname{Hom}(G_0, G)$ and hence a morphism

$$(6.4) \qquad \pi_*: L \operatorname{Hom}(G_0, F) \to \operatorname{Hom}(G_0, G) .$$

It is then easy to prove

Proposition 6.2. *The functor T_G (6.3) is cofinal if and only if π_* (6.4) is an equivalence for all G_0 in $\mathfrak{G}_0$.*

This gives us an alternative formulation of the hypotheses of Theorem 6.1. It is, however, more useful insofar as condition (6.4) is familiar for many categories $\mathfrak{G}_0$ (being sometimes true, sometimes false).

Application. Let h be a cohomology theory which satisfies the condition that the Hopf map $S^3 \to S^2$ induces the zero homomorphism in cohomology; call such a theory *good*. We have a functorial procedure [6] for putting finitely-generated coefficients G_0 into such a theory h and there results a theory $h(-;G_0)$ together with a natural universal coefficient sequence

$$(6.5) \qquad 0 \to h^n(X) \otimes G_0 \to h^n(X;G_0) \to \mathrm{Tor}\,(h^{n+1}(X), G_0) \to 0\,.$$

Thus we have a functor $W_0\colon G_0 \to \mathfrak{Cohg}$ from the category of finitely-generated Abelian groups to the category of good cohomology theories. Then $\mathfrak{G}_0$ certainly satisfies hypothesis P and $\mathfrak{Cohg}$ is cocomplete. Moreover, we verify the crucial condition of Corollary 5.2 by taking (6.5) 'to the limit', since tensor products and Tor commute with colimits over filtering categories and such colimits preserve exactness. Thus we obtain an extension $W_1\colon \mathfrak{Ab} \to \mathfrak{Cohg}$ of W_0, that is, a rule for putting any Abelian group in as a coefficient group for a good cohomology theory, and (6.5) remains true for any coefficient group G. Moreover, it is easy to show that π_*, given by (6.4), is an equivalence, so that W_1 is precisely the Kan extension of W_0.

If we wish to consider the category $\mathfrak{Coh}$ of all cohomology theories, then we should take $\mathfrak{G}_0$ to be the category of 2-torsion-free, finitely-generated Abelian groups and $\mathfrak{G}_1$ will then be the category of 2-torsion-free Abelian groups. The argument works just as well in this case.

Remarks. We would wish to work in a universe in which it is permissible to take colimits over I_G. However, the usual set-theoretical questions do not really arise in Theorem 6.1, since our hypothesis there is that I_G admits a cofinal functor from a (small) filtering category I.

References

[1] M. Artin and B. Mazur, Etale homotopy, Lecture Notes in Math. 100 (Springer, 1968).

[2] A. Deleanu and P.J. Hilton, Remark on Čech extensions of cohomology functors, Proc. Adv. Study Inst. Alg. Top. (Aarhus, Denmark, 1970).

[3] A. Deleanu and P.J. Hilton, On Kan extensions of cohomology theories and Serre classes of groups, Fund. Math. (1971) to appear.

[4] P. Gabriel and M. Zisman, Calculus of fractions and homotopy theory (Springer, 1967).
[5] J.W. Gray, Fibred and cofibred categories, Proc. La Jolla Conf. on Cat. Alg. (Springer, 1966).
[6] P.J. Hilton, Putting coefficients into a cohomology theory, Indag. Math. 32 (1970) 196–216.
[7] P.J. Hilton, On the category of direct systems and functors on groups, J. Pure and Appl. Alg. 1 (1971) 1–26.
[8] P.J. Hilton, Extensions of functors on groups and coefficients in a cohomology theory, Proc. Intern. Congr. Math., Nice (1970).

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 2 (1971) p. 219.

ERRATUM

G.M. Kelly and S. Mac Lane, Coherence in closed categories, J. Pure and Appl. Algebra 1 (1971) 97–140.

The authors' assertion in §3 that a strict monoidal functor Δ: $\underline{V} \to \underline{V}'$ induces a strict monoidal functor $\underline{N}_0(\Delta)$: $\underline{N}_0(\underline{V}) \to \underline{N}_0(\underline{V}')$ is false for a general such Δ – the suggested definition makes no sense. If, however, $\underline{V}$ and $\underline{V}'$ have the same objects, if Δ is the identity on objects, and if Δ is surjective on morphisms, all is well; and this is the only case we use in the sequel: namely when Δ is Γ: $\underline{N}(\underline{V}) \to \underline{G}$. The results of the paper are therefore quite unaffected.

There is a misprint in the statement of Lemma 5.9, which is put right by omitting everything in the second line of the lemma preceding the second "if".

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 3 (1971) pp. 221–252.

EXCISION IN ALGEBRAIC K-THEORY

Richard G. SWAN
The University of Chicago, Ill., USA

Received 12 January 1971

If I is a two-sided ideal of a ring R, the relative functor $K_0(R, I)$ depends only on I considered as a ring [1]. More precisely, if $\varphi: R \to R'$ is a ring homomorphism, I' is a two-sided ideal of R', and $\varphi: I \approx I'$, then φ induces an isomorphism $K_0(R, I) \approx K_0(R', I')$. This is an algebraic analogue of the excision property of topological K-theory. It has been an open problem for some time to determine whether the same property holds for $K_1(R, I)$. It is known to hold if φ is onto [9, Lemma 6.3]. I will give here some examples to show that excision does not hold for $K_1(R, I)$. Some explicit calculations of $K_1(R, I)$, suggested by the examples, are also given. I will also give an explicit set of generators for the kernel of $K_1(R, I) \to K_1(R', I')$ which is particularly simple in the case where R and R' are commutative. In §6, I will show that excision also fails for the functors $K_2(R, I)$ of [5, 9, 10, 16].

§ 1. An example

Let R be a ring with unit and let I be a 2-sided ideal of R. Recall [1, 9, 14] that $E(R)$ is the subgroup of $GL(R)$ generated by all elementary matrices $e_{ij}(r) = 1 + re_{ij}$ and $E(R, I)$ is the smallest normal subgroup of $E(R)$ containing all $e_{ij}(x)$ with $x \in I$. This is a normal subgroup of $GL(R, I) = \ker[GL(R) \to GL(R/I)]$. By definition, $K_1(R, I) = GL(R, I)/E(R, I)$. Clearly $GL(R, I)$ depends only on I. The following result shows that this is not the case for $K_1(R, I)$ and hence also for $E(R, I)$.

Theorem 1.1. *Let R be a ring with unit and let I be a 2-sided ideal of R. Let $R' = \mathbf{Z}1 + I$ be the smallest subring with unit containing I. If I is commutative but not central in R, then $K_1(R', I) \to K_1(R, I)$ is not an isomorphism.*

The hypothesis means that $xy = yx$ for $x, y \in I$ but $rx \neq xr$ for some $r \in R$, $x \in I$.

Proof. The element $e(r, x) = e_{21}(r)\, e_{12}(x)\, e_{21}(r)^{-1}$ for $r \in R$, $x \in I$ lies in $GL(R, I) = GL(R', I)$ and so represents an element $\epsilon(r, x) \in K_1(R', I)$. Since $e(r, x) \in E(R, I)$, $\epsilon(r, x)$ lies in the kernel of $K_1(R', I) \to K_1(R, I)$. I claim that $\epsilon(r, x) \neq 0$ if $rx \neq xr$. Since R' is commutative, the determinant $\det: K_1(R') \to U(R')$ is defined [1]. Here,

as usual, $U(R')$ is the group of units of R'. Let $\delta(r,x)$ be the image of $\epsilon(r,x)$ under the composition $K_1(R',I) \overset{\det}{\to} U(R')$. It will suffice to show that $\delta(r,x) \neq 1$. Identifying $\mathrm{GL}_2(R)$ with its image in $\mathrm{GL}(R)$, we can write

$$e(r,x) = \begin{pmatrix} 1 & 0 \\ r & 1 \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -r & 1 \end{pmatrix} = \begin{pmatrix} 1-xr & x \\ -rxr & 1+rx \end{pmatrix} .$$

Therefore $\delta(r,x) = (1+rx)(1-xr)+x(rxr)$. Since I is commutative, $(rx)x = x(rx)$ and so $(rx)(xr) = (rx)xr = x(rx)r$. Therefore $\delta(r,x) = 1+rx-xr \neq 1$.

There are, of course, many examples of rings R satisfying the conditions of Theorem 1.1. One of the simplest is obtained by taking any ring A with unit and letting

$$R = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \in M_2(A) \right\}, \quad I = \left\{ \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \in M_2(A) \right\} .$$

Here $I^2 = 0$ and it is trivial to check that I is not central. In §2, I will show that $K_1(R,I) = 0$ and $K_1(R',I) \approx A$ for this example.

The above example depends heavily on the non-commutativity of R. In §3, I will give a commutative example.

Given any example for which excision fails for K_1, we can construct a new one $(R,I) \to (R',I')$ with the additional property that $R \to R/I$ and $R' \to R'/I'$ are split epimorphisms. Let $R_1 = R \oplus I$ with $(r,i)(r',i') = (rr', ri' + ir' + ii')$ and $I_1 = 0 \oplus I$. Then $R_1 \to R$ by $(r,i) \to r+i$ is onto so $K_1(R_1, I_1) \to K_1(R,I)$ is an isomorphism. Define $R'_1 = R' \oplus I'$ and $I'_1 = 0 \oplus I'$ similarly. Then $(R_1, I_1) \to (R'_1, I'_1)$ is the required example. Note that R_1 and R'_1 are commutative if R and R' are.

Using the result of §3, we see that the following result continues to hold if all rings are required to be commutative.

Corollary 1.2. *There is no functor K_2 from rings with unit to Abelian groups such that for every Cartesian diagram*

$$\begin{array}{ccc} A & \longrightarrow & A_1 \\ \downarrow & & \downarrow{\scriptstyle f} \\ A_2 & \longrightarrow & A' \end{array}$$

of rings with unit with surjective f, there is an exact Mayer-Vietoris sequence

$$(1) \qquad K_2A_1 \oplus K_2A_2 \to K_2A' \to K_1A \to K_1A_1 \oplus K_1A_2 .$$

This continues to hold even if we restrict ourselves to diagrams where f is a split epimorphism. The result follows from [16, Prop. 9.1]. Here is a simple direct proof.

By Theorem 1.1 and the remarks above, there is a map $\varphi: (R, I) \to (R', I')$ with $\varphi: I \approx I'$ such that $K_1(R, I) \to K_1(R', I')$ is not an isomorphism and such that $R \to R/I$ and $R' \to R'/I'$ are split epimorphisms of rings. It is easy to see that the diagram

$$\begin{array}{ccc} R & \longrightarrow & R' \\ \downarrow & & \downarrow \\ R/I & \longrightarrow & R'/I' \end{array}$$

is Cartesian. If there is a Mayer-Vietoris sequence (1), we get an exact sequence

$$K_2(R') \oplus K_2(R/I) \to K_2(R'/I') \to K_1(R) \to K_1(R') \oplus K_1(R/I).$$

Since $K_2(R') \to K_2(R'/I')$ is a split epimorphism, this reduces to

(2) $$0 \to K_1(R) \to K_1(R') \oplus K_1(R/I).$$

Now let K_2' denote Milnor's K_2 and consider the exact sequence [9, 14]

$$K_2'(R) \to K_2'(R/I) \to K_1(R, I) \to K_1(R) \to K_1(R/I).$$

Since $R \to R/I$ is a split epimorphism, this sequence reduces to

$$0 \to K_1(R, I) \to K_1(R) \to K_1(R/I) \to 0\,.$$

Similarly remarks apply to R' and I' so we get a diagram

$$\begin{array}{ccccccccc} 0 & \to & K_1(R, I) & \to & K_1(R) & \to & K_1(R/I) & \to & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & K_1(R', I') & \to & K_1(R') & \to & K_1(R'/I') & \to & 0\,. \end{array}$$

Now $K_1(R, I) \to K_1(R', I')$ is onto, both groups being quotients of $\mathrm{GL}(R, I) \approx \mathrm{GL}(R', I')$. Using (2) we see that $K_1(R, I) \to K_1(R', I')$ is also injective so it is an isomorphism. This contradicts our assumption about R, I, R', I'.

Note that Milnor [9] has extended the Mayer-Vietoris sequence under the hypothesis that f *and* g are onto.

If $0 \to A \to B \to C \to 0$ is an exact sequence of rings (not necessarily with units), Gersten showed in [6] that there is an exact sequence

$$K_1(A) \to K_1(B) \to K_1(C) \to K_0(A) \to K_0(B) \to K_0(C)\,.$$

The following result also holds if all rings are required to be commutative, using the results of §3. It says that Gersten's sequence cannot be extended to any K_2.

Corollary 1.3. *There is no functor K_2 from rings to Abelian groups such that for any exact sequence $0 \to A \to B \to C \to 0$ of rings, there is an exact sequence*

$$K_2B \to K_2C \to K_1A \to K_1B. \tag{4}$$

This continues to hold even if we require B and C to have units and require $B \to C$ to be a split epimorphism. In this case (4) would reduce to $0 \to K_1A \to K_1B$. For the proof consider the same R, I, R', I' used in the proof of Corollary 1.2. If there is an exact sequence (4), we get a diagram

$$\begin{array}{ccccccccc} 0 & \to & K_1(I) & \to & K_1(R) & \to & K_1(R/I) & \to & 0 \\ & & \approx\downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & K_1(I') & \to & K_1(R') & \to & K_1(R'/I') & \to & 0\,. \end{array}$$

From this we deduce easily the exact sequence (2). The argument used to deduce Corollary 1.2 now applies.

§ 2. Radical ideals

An ideal I (left, right, or 2-sided) of a ring R (not necessarily with unit) is called a radical ideal of R if it lies in the Jacobson radical of R. If R is equal to its own radical, it is called a radical ring. It is a standard fact [7] that I is a radical ideal if and only if for each $x \in I$ there is some $y \in R$ with $x + y + xy = 0 = x + y + yx$. Now $y = -x - xy = -x - yx$ so $y \in I$. It follows that I is a radical ring. Conversely, if I is a radical ring and $x \in I$, we can find $y \in I \subset R$ as above so I is a radical ideal. Therefore an ideal I of R is a radical ideal if and only if I is a radical ring. In particular, the property of being a radical ideal is invariant under excision.

In this section, I will compute $K_1(R, I)$ for the case where I is a radical ideal. The result is closely related to some recent work of Wasserstein [18]. However, I will not use his methods except for his Lemma 1.1. Instead I will give a proof similar to Dieudonne's construction of determinants [4]. If R is commutative, the ordinary determinant can be used. However, in this case, the result is already known [3, Lemma 3.2].

As usual, $U(R)$ will denote the group of units of R and $U(R, I)$ is defined to be the kernel of $U(R) \to U(R/I)$. Clearly $U(R, I)$ depends only on I.

Theorem 2.1. *Let R be a ring with unit and let I be a 2-sided radical ideal of R. Then $K_0(R, I) = 0$ and $K_1(R, I) = U(R, I)/W(R, I)$ where $W(R, I)$ is the subgroup of $U(R, I)$ generated by all elements of the form $(1 + rx)(1 + xr)^{-1}$ with $r \in R$, $x \in I$.*

The group $W(R, I)$ was introduced by Wasserstein [18] and was also considered by Silvester in [11].

Proof. Let I^+ be the result of formally adjoining a unit to I. Then $0 \to I \to I^+ \to \mathbf{Z} \to 0$ is a split exact sequence. The exact sequence

$$K_1(I^+) \to K_1(\mathbf{Z}) \to K_0(I^+, I) \to K_0(I^+) \to K_0(\mathbf{Z})$$

therefore reduces to a split exact sequence

$$0 \to K_0(I^+, I) \to K_0(I^+) \to K_0(\mathbf{Z}) \to 0\,.$$

If P is a projective I^+-module, then P/IP is projective over $\mathbf{Z}$ and therefore free. Suppose P is finitely generated. Since I is a radical ideal of I^+ by the remarks at the beginning of this section, we can apply the usual projective cover argument [14, p. 89] to conclude that P is free. Therefore $K_0(I^+) = \mathbf{Z}$ and it follows that $K_0(I^+, I) = 0$. Since excision holds for K_0, we see that $K_0(R, I) = 0$ using the map $I^+ \to R$ which is the identity on I and sends 1 to 1.

We now consider K_1. Suppose $a \in U(R, I)$ and $r \in U(R)$. Following the argument of [18, Cor. 1.3] we set $x = (a - 1)r^{-1} \in I$ and conclude that $(1 + rx)(1 + xr)^{-1} = rar^{-1}a^{-1}$. It follows that $[U(R), U(R, I)] \subset W(R, I)$. In particular $U(R, I)/W(R, I)$ is Abelian. Since $U(R, I) = \mathrm{GL}_1(R, I) \subset \mathrm{GL}(R, I)$ we have a map $U(R, I) \to K_1(R, I)$. Wasserstein [18, Lemma 1.1] has shown that $W(R, I)$ lies in the kernel of this so we have a map $\varphi\colon U(R, I)/W(R, I) \to K_1(R, I)$.

Lemma 2.2. *φ is onto.*

Proof. This is a well-known result [1, V Prop. 3.4, Th. 4.2]. Let $A \in \mathrm{GL}(R, I)$. Since I is a radical ideal, the diagonal elements of A are units of R. All other elements of A lie in I. Therefore, by multiplying on the left by elements of the form $e_{ij}(x)$ with $x \in I$, we can reduce X to diagonal form. Since $\left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right) \in E(R, I)$ for $a \in U(I)$, we can reduce A to the form $A' = \mathrm{diag}(a, 1, 1, \ldots)$ by multiplying on the left by elements of $E(R, I)$. Clearly $A' \in \mathrm{GL}_1(R, I)$.

To show that φ is injective, we construct a determinant $\delta\colon \mathrm{GL}(R, I) \to U(R, I)/W(R, I)$. If $x \in U(R, I)$, let $\bar{x}$ denote its image in $U(R, I)/W(R, I)$. Define $\delta_1\colon \mathrm{GL}_1(R, I) \to$ $\to U(R, I)/W(R, I)$ by $\delta_1(x) = \bar{x}$. If $X \in \mathrm{GL}_n(R, I)$ for $n \geq 2$, let

$$Y = \left(\prod_{i=2}^{n} e_{i1}(-x_{i1}x_{11}^{-1})\right) X = \begin{pmatrix} x_1 & * \cdots * \\ 0 & \\ \vdots & Z \\ 0 & \end{pmatrix}$$

and set $\delta_n(X) = \bar{x}_{11}\delta_{n-1}(Z)$.

Lemma 2.3. (a) *If $x \in I$ then $\delta_n(e_{ij}(x)X) = \delta_n(X)$.*

(b) *If $a \in U(I)$ and X' is obtained from X by multiplying one row on the left by a, then $\delta_n(X') = \bar{a}\delta_n(X)$.*

Proof. Property (b) follows immediately by induction on n. The same is true of (a) provided that $i \neq 1$ so that $e_{ij}(x)X$ and X have the same first row. Suppose now that $i = 1$. Let

$$U = \left(\prod_{k \neq 1, j} e_{k1}(-x_{k1}x_{11}^{-1}) \right) X = \begin{pmatrix} x_{11} & * \cdots * \\ 0 & \\ 0 & \\ x_{j1} & * \\ 0 & \\ \vdots & \end{pmatrix}.$$

Since $e_{1j}(x)e_{k1}(t) = e_{kj}(-tx)e_{k1}(t)e_{1j}(x)$ by the Steinberg relations [9, 14], we can write $e_{1j}(x)X = Ee_{1j}(x)U$ where E is a product of elements $e_{kl}(s)$ with $k \neq 1$. It follows from the part of (a) already proved that $\delta_n(e_{1j}(x)X) = \delta_n(Ee_{1j}(x)U) = \delta_n(e_{1j}(x)U)$. Similarly $\delta_n(X) = \delta_n(U)$. Let A_i be the ith row of U. Then the rows of $e_{1j}(x)U$ are $A_1 + xA_j, A_2, \ldots, A_n$. To compute $\delta_n(U)$ we write

$$U' = e_{j1}(\lambda)U = \begin{pmatrix} u_{11} & * \cdots * \\ 0 & \\ \vdots & \\ 0 & V \end{pmatrix}$$

where $u_{j1} + \lambda u_{11} = 0$ and get $\delta_n(U) = \bar{u}_{11}\delta_{n-1}(V)$. The rows of U' are $A_1, \ldots, A_j + \lambda A_1, \ldots$. To compute $\delta_n(e_{1j}(x)U)$ we write

$$U'' = e_{j1}(\mu)e_{1j}(x)U = \begin{pmatrix} u_{11} + xu_{j1} & * \cdots * \\ 0 & \\ \vdots & V' \\ 0 & \end{pmatrix}$$

where $u_{j1} + \mu(u_{11} + xu_{j1}) = 0$ and get $\delta_n(U) = (\overline{u_{11} + xu_{j1}})\delta_{n-1}(V')$. The rows of U'' are $A_1 + xA_j, A_2, \ldots, A_j + \mu(A_1 + xA_j), \ldots$.

Now $(1 + \mu x)u_{j1} + \mu u_{11} = 0$. Since $x \in I$, $1 + \mu x \in U(R, I)$ so $u_{j1} = -(1 + \mu x)^{-1}\mu u_{11}$. But $u_{j1} = -\lambda u_{11}$ and $u_{11} \in U(R, I)$ so $\lambda = (1 + \mu x)^{-1}\mu$. Therefore $A_j + \mu(A_1 + xA_j) = (1 + \mu x)A_j + \mu A_1 = (1 + \mu x)(A_j + \lambda A_1)$. Thus V' differs from V only in having its $(j-1)$st row multiplied on the left by $1 + \mu x$. By (b), $\delta_{n-1}(V') = (\overline{1 + \mu x})\delta_{n-1}(V)$. Therefore $\delta_n(e_{1j}(x)U) = \alpha\delta_n(U)$ where $\alpha = (\overline{u_{11} + xu_{j1}})(\overline{1 + \mu x})\bar{u}_{11}^{-1} = (\overline{1 + xu_{j1}u_{11}^{-1}})(\overline{1 + \mu x})$ since $U(R, I)/W(R, I)$ is commutative. Now $(1 + \mu x)u_{j1} + \mu u_{11} = 0$ so $u_{j1}u_{11}^{-1} = (1 + \mu x)^{-1}\mu$. Therefore $(1 + xu_{j1}u_{11}^{-1})(1 + \mu x) = (1 - x(1 + \mu x)^{-1}\mu)(1 + \mu x) \equiv (1 - \mu x(1 + \mu x)^{-1})(1 + \mu x) = 1 \bmod W(R, I)$, so $\alpha = 1$.

Corollary 2.4. (a) *If $X, Y \in \mathrm{GL}_n(R, I)$, then $\delta_n(XY) = \delta_n(X)\delta_n(Y)$.*

(b) *If $X \in \mathrm{GL}_n(R, I)$ has the form $X = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ with $A \in \mathrm{GL}_p(R, I)$, $C \in \mathrm{GL}_q(R, I)$, then $\delta_n(X) = \delta_p(A)\delta_q(C)$.*

Proof. For (a), write $X = ED$ as in the proof of Lemma 2.2 where D is diagonal and E is a product of elements $e_{ij}(x)$ with $x \in I$. Then $XY = EDY$, $X = ED$, and the result follows from Lemma 2.3(a) and (b). Property (b) is immediate from the definition of δ_n by induction on n.

It follows from (b) that the δ_n define a homomorphism $\delta: \mathrm{GL}(R, I) \to U(R, I)/W(R, I)$. Clearly all $e_{ij}(x)$ with $x \in I$ lie in the kernel of δ. To show that $E(R, I) \subset \ker \delta$, we must show that $\ker \delta$ is normalized by $E(R)$.

Lemma 2.5. *If $Y \in E(R)$ and $X \in \mathrm{GL}(R, I)$, then $\delta(YXY^{-1}) = \delta(X)$.*

Proof. It will suffice to do the case $Y = e_{ij}(r)$ with $r \in R$. As in the proof of Lemma 2.2, X is a product of elements $e_{kl}(x)$ with $x \in I$ and a diagonal matrix D so it is enough to do the cases $X = e_{kl}(x)$, $x \in I$ and $X = D$. If $X = e_{ij}(x)$ or $e_{kl}(x)$ with $k \neq j$, $l \neq i$, then $YXY^{-1} = X$. If $X = e_{jk}(x)$, $k \neq i$, then $YXY^{-1} = e_{ik}(rx)e_{jk}(x)$ so $\delta(YXY^{-1}) = 1 = \delta(X)$. If $X = \mathrm{diag}(d_1, d_2, \ldots)$ then $XY^{-1}X^{-1} = e_{ij}(-d_i r d_j^{-1})$ so $YXY^{-1} = e_{ij}(r)\, e_{ij}(-d_i r d_j^{-1})X$ and we use Lemma 2.3(a).

In the only remaining case $X = e_{ji}(x)$. Omitting all but the ith and jth rows and columns from the notation we can write

$$X = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & r \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad YXY^{-1} = \begin{pmatrix} 1+rx & -rxr \\ x & 1-xr \end{pmatrix}.$$

Applying the definition of δ, we see that $\delta(YXY^{-1})$ is the image of $(1+rx)(1-xr+x(1+rx)^{-1}rxr) = (1+rx)(1-x(1+rx)^{-1}r)$ using the fact that $(1+rx)^{-1}rx = 1-(1+rx)^{-1}$. Let $y = x(1+rx)^{-1}$. Then $1-ry = 1-rx(1+rx)^{-1} = (1+rx)^{-1}$ so $\delta(YXY^{-1}) = (1-ry)^{-1}(1-yr) \in W(R, I)$.

Therefore, δ defines a map $\delta: K_1(R, I) \to U(R, I)/W(R, I)$. The definition of δ_1 shows that $\delta\varphi = \mathrm{id}$ so φ is an isomorphism. This proves Theorem 2.1.

Corollary 2.6. *If $I^2 = 0$, then $K_1(R, I) \approx I/W$ where W is the subgroup of I generated by all $rx - xr$ for $r \in R$, $x \in I$.*

In fact, if $I^2 = 0$, then $U(R, I) \approx I$ where $a \in U(R, I)$, corresponds to $a - 1 \in I$. In particular, for the 2×2 triangular matrix example of §1 we have $K_1(R, I) = 0$ while $K_1(R', I) \approx I$.

If R in Theorem 2.1 is commutative, then $W(R, I) = 0$ so $K_1(R, I) = U(R, I)$. This was already proved in [3, Lemma 3.2]. It follows that excision holds if we consider only commutative rings and radical ideals. In particular, if

$$\begin{tikzcd} A \arrow[r] \arrow[d] & A_1 \arrow[d, "f"] \\ A_2 \arrow[r] & A' \end{tikzcd}$$

is a Cartesian diagram of commutative rings, f is onto, and $I = \ker f$ is a radical ideal, the usual ladder argument [14] gives us the Mayer-Vietoris sequence

$$K_2(A_1) \oplus K_2(A_2) \to K_2(A') \to K_1(A) \to K_1(A_1) \oplus K_1(A_2) \to \dots$$

using Milnor's K_2. Note that $J = \ker[A \to A_2]$ is also a radical ideal by the remarks at the beginning of this section. In order to put $K_2(A)$ at the left, we would have to show that $K_2(A, J) \to K_2(A_1, I)$ is onto but this is false in general (see §6).

Theorem 2.1 is easily extended to the case where R has no unit. In this case we have, practically be definition, $K_i(R, I) = K_i(R^+, I)$ and $U(R, I) = U(R^+, I)$. Clearly $U(R^+, I) = U(I^+, I) = U(I)$ depends only on I.

Corollary 2.7. *If I is a 2-sided radical ideal of R, then $K_0(R, I) = 0$ and $K_1(R, I) = U(I)/[U(I), U(I)]\, W(R, I)$ where $W(R, I)$ is the subgroup of $U(I)$ generated by all elements of the form $(1+rx)(1+xr)^{-1}$ for $x \in I$, $r \in R$.*

Expressions of the form $1 + rx$ make sense since we view $U(I)$ as a subgroup of $U(I^+)$ or $U(R^+)$

Proof. Using Theorem 2.1 we need only show that $W(R^+, I) = [U(I), U(I)]\, W(R, I)$. Clearly $W(R^+, I) \supset [U(I), U(I)]\, W(R, I)$. Now $W(R^+, I)$ is generated by elements $w = (1 + (n+r)x)(1 + x(n+r))^{-1}$ where $n \in Z$, $r \in R$. But $w = (1+ry)(1+nx)(1+yr)^{-1} \times (1+nx)^{-1}$ where $y = (1+nx)^{-1}x$. Modulo $[U(I), U(I)]$ we can change the order of the factors so $w \equiv (1+ry)(1+yr)^{-1} \in W(R, I)$.

In particular, if R is a radical ring, then $K_0(R) = 0$ and $K_1(R) = U(R)/[U(R), U(R)]\, W(R)$ where $W(R)$ is the subgroup of $U(R)$ generated by all $(1+xy)(1+yx)^{-1}$ with $x, y \in R$.

§3. A commutative example

If R is a commutative ring with unit and I is an ideal of R, the map $U(R, I) \to K_1(R, I)$ given by $U(R, I) = \mathrm{GL}_1(R, I) \subset \mathrm{GL}(R, I)$, is split by the determinant map $\det: K_1(R, I) \to U(R, I)$. We write $K_1(R, I) = U(R, I) \oplus \mathrm{SK}_1(R, I)$, a natural decomposition [1]. Since $U(R, I)$ depends only on I, it is sufficient to look at $\mathrm{SK}_1(R, I)$ in considering the question of excision.

As in [1, Ch. VI], we define W_I to be the set of all (a, b) with $Ra + Rb = R$, $a \equiv 1 \pmod I$ and $b \equiv 0 \pmod I$. For $(a, b) \in W_I$, we can find $c, d \in R$ such that $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(R, I)$ and the image $\left[\begin{smallmatrix} b \\ a \end{smallmatrix}\right]$ of this in $\mathrm{SK}_1(R, I)$ is independent of the choice of a and b [1, Ch. VI]. The Mennicke symbol $\left[\begin{smallmatrix} b \\ a \end{smallmatrix}\right]$ has the following properties [1, Ch. VI]. I will write $\mathrm{SK}_1(R, I)$ multiplicatively here:

(MS1a) $\quad \begin{bmatrix} b+ta \\ a \end{bmatrix} = \begin{bmatrix} b \\ a \end{bmatrix} \quad$ for $t \in I$,

(MS1b) $$\begin{bmatrix} b \\ a+tb \end{bmatrix} = \begin{bmatrix} b \\ a \end{bmatrix} \quad \text{for } t \in R\,,$$

(MS2a) $$\begin{bmatrix} b_1 \\ a \end{bmatrix}\begin{bmatrix} b_2 \\ a \end{bmatrix} = \begin{bmatrix} b_1 b_2 \\ a \end{bmatrix}.$$

(MS2b) $$\begin{bmatrix} b \\ a_1 \end{bmatrix}\begin{bmatrix} b \\ a_2 \end{bmatrix} = \begin{bmatrix} b \\ a_1 a_2 \end{bmatrix}.$$

Remark. As an illustration of the useful technique of universal examples, I will show how to deduce the relations (MS2) from the corresponding relations in the absolute case $I = R$. Since I will only consider the canonical Mennicke symbol with values in $\mathrm{SK}_1(R, I)$, this result does not supplant Proposition 1.7 of [1]. In the absolute case, it follows from (MS1) [1] that $[{b \atop a}] = [{b+a \atop a}] = [{b+a \atop -b}] = [{a \atop -b}]$. Assuming (MS2a) or (MS2b) we deduce easily [1] that $[{a \atop -1}] = [{0 \atop -1}] = 1$ so $[{b \atop a}] = [{a \atop b}]$. Thus either of (MS2a) or (MS2b) implies the other. To prove (MS2b) in the relative case, consider the ring $A = \boldsymbol{Z}[x_1, x_2, y, u_1, u_2, v_1, v_2]/(u_1 x_1 - v_1 y - 1, u_2 x_2 - v_2 y - 1)$ and the ideal $J = (x_1 - 1, x_2 - 1, y)$. If (a_1, b) and $(a_2, b) \in W_I$, choose elements p_1, p_2, q_1, q_2 with $p_i a_i + q_i b = 1$ and define a map $A \to R$ by sending x_i to a_i, y to b, u_i to p_i, and v_i to q_i. This sends J into I. Therefore it will suffice to show that $[{y \atop x_1 x_2}] = [{y \atop x_1}][{y \atop x_2}]$ in $\mathrm{SK}_1(A, J)$. Now $A/J = \boldsymbol{Z}[v_1, v_2]$ so the sequence $0 \to J \to A \to A/J \to 0$ splits. Therefore, the sequence

$$K_2(A) \to K_2(A/J) \to K_1(A, J) \to K_1(A) \to \dots$$

shows that $K_1(A, J) \to K_1(A)$ is a monomorphism so it is enough to check the relation in $K_1(A)$ i.e., for the absolute case. The proof of (MS2a) is similar.

If R is a commutative Noetherian domain with unit, and $\dim R \leqslant 1$ then [1, VI, Th. 2.3] asserts that $\mathrm{SK}_1(R, I)$ is generated by the $[{b \atop a}]$ with $(a, b) \in W_I$ and (MS1), (MS2) are a set of defining relations for $\mathrm{SK}_1(R, I)$. Note that the ring R enters only in the relations (MS1b) since W_I and the other relations depend only on I.

Let p be an odd prime and let ζ be a primitive pth root of 1. Let $K = \boldsymbol{Q}(\zeta)$ and let $A = \boldsymbol{Z}[\zeta]$ be the ring of integers of K. Let $I = pA$ and $R = \boldsymbol{Z} + I \subset A$.

Theorem 3.1. $\mathrm{SK}_1(R, I) \to \mathrm{SK}_1(A, I)$ *is not an isomorphism.*

Proof. By [2, Th. 3.6] we have $\mathrm{SK}_1(A, I) = 0$. I will use the methods of [2] to show that $\mathrm{SK}_1(R, I) \neq 0$. I will actually prove a slightly more general result which will be useful in §6. Let $\lambda = 1 - \zeta$. Then $\mathfrak{p} = (\lambda)$ is prime and $(p) = \mathfrak{p}^{p-1}$. Let $B = \boldsymbol{Z} + \mathfrak{p}^2 \subset A$. As usual, μ_p denotes the group generated by ζ.

Lemma 3.2. *There is a non-trivial homomorphism* $\mathrm{SK}_1(B, I) \to \mu_p$ *given by sending* $[{b \atop a}]$ *to* $(\frac{b}{a})_p$.

Proof. The Jacobi symbol $(\frac{b}{a})_p$ here is the one considered in [2]. This satisfies (MS1a), (MS2a), and (MS2b). We must verify (MS1b). As in the proof of Proposition 3.1 of [2], we have

$$\left(\frac{b}{a}\right)_p = \prod_{\mathfrak{p}|b,\ \mathfrak{p}\nmid p} \left(\frac{a}{\mathfrak{p}}\right)_p^{\operatorname{ord}_{\mathfrak{p}} b} \prod_{\mathfrak{p}|p} \left(\frac{a,b}{\mathfrak{p}}\right)_p$$

and it is sufficient to show that the second term is unchanged if we replace a by $a+tb$ with $t\in B$. Since there is only one prime $\mathfrak{p}$ in K, the term in question is just $(\frac{a,b}{\mathfrak{p}})_p$ which we abbreviate to (a,b). Now $e = \operatorname{ord}_{\mathfrak{p}}(p) = p-1$. Since $a\equiv 1 \bmod pR$, we see that $a\in U_{\mathfrak{p}}(p-1)$. By [2, A17] we have $(c,b) = 1$ if $c\equiv 1 \bmod \mathfrak{p}^{p+1}$. If $t\in \mathfrak{p}^2$, we have $(a+tb,b) = (a,b)(1+a^{-1}tb,b)$ but $1+a^{-1}tb\equiv 1 \bmod \mathfrak{p}^{p+1}$. The set of $t\in B$ such that $(a+tb,b) = (a,b)$ for all $a\equiv 1$, $b\equiv 0 \bmod I$ is obviously a subgroup of B. We have just seen that it contains $\mathfrak{p}^2$ so it will suffice to show that it contains -1. Now $(a-b,b) = (a,b)(1-a^{-1}b,b)$. But $a\equiv 1 \bmod p$ so $a^{-1}\equiv 1 \bmod p$. Let $a^{-1} = 1+cp$. Then $(1-a^{-1}b,b) = (1-b-cpb,b) = (1-b,b)(1-(1-b)^{-1}cpb,b)$. The second term is 1 by [2, A17] as above because $pb\equiv 1 \bmod p^2$ and so mod $\mathfrak{p}^{p+1}$. But $(1-b,b) = 1$ for any b [2, A13].

It remains to show that the resulting map $SK_1(B,I)\to\mu_p$ is non-trivial. By [2, A11] we can find $a\in A$ such that $a\equiv 1 \bmod p$ and Aa is prime. Let $c\in A$ map onto a generator of $U(A/Aa)$. Then $(\frac{c}{a})_p \neq 1$ so at least one of $(\frac{p}{a})_p$ and $(\frac{pc}{a})_p$ is non-trivial. Clearly (a,p) and (a,pc) lie in W_I. Since $SK_1(R,I)\to SK_1(B,I)$ is onto, it follows that $SK_1(R,I)\neq 0$.

A more explicit example is obtained by letting $p=3$, $a=1+p\zeta$ and $b=p$. Here Aa is prime since $N(a) = 13$. By definition,

$$\left(\frac{b}{a}\right)_p \equiv b^{(Na-1)/p} = 3^4 \equiv 3 \bmod 13\ ,$$

so $[\frac{3}{1+3\zeta}]\neq 1$ in $SK_1(R',I)$. By (MS1b), the image of this in $SK_1(R,I)$ is 1.

It would be very interesting to have commutative examples of characteristic p for $p\neq 0$. In §5, I will show that the methods of the present section will not yield such examples. The problem is discussed further at the end of §4.

The proof of Lemma 3.2 obviously extends to more general cases. For example, let K be a totally imaginary algebraic number field, let A be the ring of integers of K, and let p be a (rational) prime such that $\mu_p\subset K$. Let $(p) = \prod \mathfrak{p}_i^{e_i}$ in A. Since $\mathbf{Z}[\zeta]\subset A$, we see that $(p-1)|e_i$. Write $e_i = (p-1)e_i'$. Let $I = \prod \mathfrak{p}_i^{n_i}$ and $B = \mathbf{Z} + \prod \mathfrak{p}_i^{m_i}$. Then $I\subset B$ if $n_i\geqslant m_i$. If, in addition, $m_i+n_i\geqslant 1+pe_i'$, the proof of Lemma 3.1 shows that there is a non-trivial homomorphism $SK_1(B,I)\to\mu_p$ which sends $[\frac{b}{a}]$ to $(\frac{b}{a})_p$. If also $n_i < pe_i'$ for some i, then $SK_1(A,I) = 0$ by [2, Th. 3.6].

If $p\neq 2$, we can choose $I = pA$ so $n_i = e_i$ and let $m_i = 1+e_i'$. If $p=2$, $I=pA$ will not work since $e_i' = e_i$ but we can let $n_i = 1+e_i$, $m_i = e_i$. For this choice of n_i, $SK_1(A,I)$ will be zero provided that 2 ramifies in K so that some $e_i\geqslant 2$.

§4. Generators for the kernel

In this section, I will show how to find a set of generators for the kernel of an excision map. The result is particularly simple in the commutative case (see Theorem 4.6).

Definition. Let I be a 2-sided ideal of a ring R with unit. For $i \neq j$, $r_1, \ldots, r_n \in R$ and $x \in I$, define $e_{ij}(r_1, \ldots, r_n; x)$ to be $e_{ij}(x)$ if $n = 0$ and, for $n > 0$,

(1) $$e_{ij}(r_1, \ldots, r_n; x) = e_{ij}(r_1)\, e_{ji}(r_2, \ldots, r_n; x)\, e_{ij}(r_1)^{-1}.$$

Theorem 4.1. *The group $E(R, I)$ is generated by the elements $e_{ij}(r_1, \ldots, r_n; x)$ for all n, $i \neq j$, $r_1, \ldots, r_n \in R$, $x \in I$.*

The proof uses only the Steinberg relations and so can be carried out in the Steinberg group $\mathrm{St}(R)$. I will give a more general version of Theorem 4.1 which also applies to all Chevalley group schemes and their Steinberg groups [13]. For any ring R with unit $\mathrm{St}(R)$ has generators $x_{ij}(r)$ for $i \neq j$, $r \in R$. The Steinberg relations are

(S1) $$x_{ij}(r+s) = x_{ij}(r)\, x_{ij}(s),$$

(S2) $$[x_{ij}(r), x_{kl}(s)] = 1 \quad \text{if} \quad i \neq l,\ j \neq k,$$

(S3) $$[x_{ij}(r), x_{jk}(s)] = x_{ik}(rs) \quad \text{if} \quad i \neq k,$$

(S4) $$[x_{jk}(r), x_{ij}(s)] = x_{ik}(-sr) \quad \text{if} \quad i \neq k.$$

The relations (S3) and (S4) are equivalent by taking inverses. The map $\varphi\colon \mathrm{St}(R) \to E(R)$ is defined by $\varphi(x_{ij}(r)) = e_{ij}(r)$ and $E(R, I) = \varphi(\mathrm{St}'(R, I))$ where $\mathrm{St}'(R, I)$ is the smallest normal subgroup of $\mathrm{St}(R)$ which contains all $x_{ij}(x)$ with $x \in I$.

Suppose, more generally, that we have a group G generated by elements $x_\alpha(r)$ for $r \in R$ and $\alpha \in \Sigma$ where Σ is an index set. We assume that Σ has an involution which we denote by $\alpha \mapsto -\alpha$ with $\alpha \neq -\alpha$, $-(-\alpha) = \alpha$. The elements $x_\alpha(r)$ are assumed to satisfy the following relations (and possibly others):

(a) $$x_\alpha(r+s) = x_\alpha(r)\, x_\alpha(s).$$

(b) $$\text{If } \alpha \neq -\beta,\ [x_\alpha(r), x_\beta(s)] = \prod x_{\gamma_i}(t_i)$$

where the γ_i lie in a certain finite subset $S_{\alpha\beta}$ of Σ. We make the following additional hypotheses

(c) $$\text{If } \gamma \in \{\beta\} \cup S_{\alpha\beta} \quad \text{then} \quad -\gamma \notin \{\beta\} \cup S_{\alpha\beta}.$$

(d) The set $\{\beta\} \cup S_{\alpha\beta}$ can be ordered in such a way that for all $\gamma, \delta \in \{\beta\} \cup S_{\alpha\beta}$ we have $S_{\gamma\delta} \subset \{\zeta \in S_{\alpha\beta} \mid \zeta > \gamma, \zeta > \delta\}$.

(e) In relation (b) if $r \in I$ or $s \in I$, then all $t_i \in I$.

Condition (c) is needed to insure the existence of $S_{\gamma\delta}$ in (d).

Example 1. Let $G = \mathrm{St}(R)$ or $E(R)$ and let Σ be the set of pairs (i, j) with $i \neq j$. Define $-(i, j) = (j, i)$. Then (a) is (S1) and (b) is the set of relations (S2), (S3), (S4). The set $S_{\alpha\beta}$ has at most one element so there is no difficulty in verifying (c), (d), and (e). In (d) we order $\{\beta\} \cup S_{\alpha\beta}$ by choosing β to be least.

Example 2. Let G be the Steinberg group associated with a Chevalley group scheme [13]. Let Σ be the root system. Then $\{\beta\} \cup S_{\alpha\beta}$ consists of all roots of the form $p\alpha + q\beta$ where $p \geq 0$, $q > 0$ are integers. If $\alpha = \beta$, $S_{\alpha\beta}$ is empty. If $\alpha \neq \pm\beta$ then α, β are linearly independent and we order $S_{\alpha\beta}$ by ordering the pairs (q, p) lexicographically. Conditions (c), (d), and (e) are clearly satisfied.

In the general situation considered above define $x_\alpha(r_1, \dots, r_n; x)$ for $r_1, \dots, r_n \in R$, $x \in I$, to be $x_\alpha(x)$ if $n = 0$ and, for $n > 0$,

$$x_\alpha(r_1, \dots, r_n; x) = x_\alpha(r_1)\, x_{-\alpha}(r_2, \dots, r_n; x)\, x_\alpha(r_1)^{-1}.$$

Let N be the smallest normal subgroup of G containing all $x_\alpha(x)$ for $\alpha \in \Sigma$, $x \in I$.

Theorem 4.2. *The group N is generated by the elements $x_\alpha(r_1, \dots, r_n; x)$ for all n, $\alpha \in \Sigma$, $r_1, \dots, r_n \in R$, $x \in I$.*

Proof. The elements clearly lie in N. Let H be the subgroup they generate. Since all $x_\alpha(x)$ with $x \in I$ lie in H, it will suffice to show that H is normal in G. For this, it is enough to show that the generators $x_\alpha(r)$ of G normalize H. The following lemma gives a more precise statement of this.

Lemma 4.3. *If $\alpha \neq -\beta$, then $x_\alpha(r)\, x_\beta(r_1, \dots, r_n; a)\, x_\alpha(r)^{-1}$ is a product of elements of the form $x_\gamma(s_1, \dots, s_m; b)$ with $\gamma \in \Sigma$, $s_1, \dots, s_m \in R$, $b \in I$, and $m \leq n$.*

If $\alpha = -\beta$, our element will just be $x_\alpha(r, r_1, \dots, r_n; a)$. Therefore, in all cases, we see that conjugation by $x_\alpha(r)$ sends the generators of H into elements of H.

To prove Lemma 4.3, we use induction on n. If $n = 0$, the result follows from relation (b) together with (e). Suppose now that $k \geq 1$. Then

$$x_\alpha(r)\, x_\beta(r_1, \dots, r_k; a)\, x_\alpha(r)^{-1} = x\, x_{-\beta}(r_2, \dots, r_k; a)\, x^{-1} ,$$

where

$$x = x_\alpha(r)\, x_\beta(r_1) = \prod x_{\gamma_i}(t_i) x_\beta(r_1)$$

by relation (b). Under the hypothesis that Lemma 4.3 holds for $n < k$, I will prove the following more general statement.

(*) *Let $\alpha, \beta \in \Sigma$ with $\alpha \neq -\beta$. Let $S \subset \{\beta\} \cup S_{\alpha\beta}$ and let $x = \prod x_{\gamma_i}(t_i)$ where each $\gamma_i \in S$ and $t_i \in R$. Let $\delta \in \Sigma$, $r_1, \ldots, r_n \in R$, $a \in I$, and $n < k$. Then $x\, x_\delta(r_1, \ldots, r_n; a) x^{-1}$ is a product of elements of the form $x_\epsilon(s_1, \ldots, s_m, b)$ with $\epsilon \in \Sigma$ and $m \leqslant n+1$.*

Proof. Let $S = \{\sigma_1, \ldots, \sigma_s\}$ where $\sigma_1 < \ldots < \sigma_s$ in the ordering of (d). If $T \subset \{\beta\} \cup S_{\alpha\beta}$, write $T = \{\tau_1, \ldots, \tau_t\}$ where $\tau_1 < \tau_2 < \ldots < \tau_t$. We write $S < T$ if $\sigma_1 < \tau_1$ or if $\sigma_1 = \tau_1$, $\sigma_2 < \tau_2$ (a partial ordering).

If $\sigma_\nu \neq -\delta$, move all terms $x_{\gamma_i}(t_i)$ in x with $\gamma_i = \sigma_\nu$ to the right and combine them using (a). By (b) this introduces new factors $x_{\zeta_j}(u_j)$ into x but by (d) all such ζ_j satisfy $\zeta_j > \sigma_\nu$ and lie in $S_{\alpha\beta}$. By the induction hypothesis $x_{\sigma_\nu}(r)\, x_\delta(r_1, \ldots, r_n; a)\, x_{\sigma_\nu}(r)^{-1}$ is a product of terms of the form $x_\epsilon(s_1, \ldots, s_m, b)$ with $m \leqslant n$. We must conjugate these by the element y where $x = y x_{\sigma_\nu}(r)$ is the expression obtained by moving all $x_{\gamma_i}(t_i)$ with $\gamma_i = \sigma_\nu$ to the right.

Now $y = \prod x_{\eta_i}(v_i)$ where all η_i lie in a set $T \subset \{\beta\} \cup S_{\alpha\beta}$. This set differs from S in having σ_ν missing and in having (possibly) new elements $\zeta_j > \sigma_\nu$. Since $\sigma_1 < \sigma_2$ we can always choose $\nu = 1$ or 2 with $\sigma_\nu \neq -\delta$. It follows that $T > S$. Since $\{\beta\} \cup S_{\alpha\beta}$ is finite we can only repeat the argument a finite number of times. The only cases where we cannot repeat the argument are those where $x = 1$ or $x = x_\sigma(t)$ with $\sigma = -\delta$. In the latter case, our element is $x_\sigma(t, r_1, \ldots, r_n; a)$. This completes the proof.

We now consider the problem of excision. Let $f\colon A \to B$ be a homomorphism of rings with unit. Let $I \subset A$, $J \subset B$ be 2-sided ideals such that $f\colon I \approx J$. We wish to determine the kernel of the epimorphism $K_1(A, I) \to K_1(B, J)$. Since $f\colon A \to f(A)$ is onto, we see that $K_1(A, I) \to K_1(f(A), J)$ is an isomorphism. Therefore, there is no loss of generality if we assume that f is an inclusion. Therefore we suppose that A is a subring of B (with the same unit) and $I \subset A$ is an ideal of B. Then $K_1(B, I) = \mathrm{GL}(B, I)/E(B, I)$, and $K_1(A, I) = \mathrm{GL}(A, I)/E(A, I)$. Since $\mathrm{GL}(A, I) = \mathrm{GL}(B, I)$, we get an exact sequence

$$0 \to E(B, I)/E(A, I) \to K_1(A, I) \to K_1(B, I) \to 0\,.$$

Using Theorem 4.1, we see that the kernel of $K_1(A, I) \to K_1(B, I)$ is generated by the images in $K_1(A, I)$ of the elements $e_{ij}(b_1, \ldots, b_n; x)$ where $b_i \in B$, $x \in I$.

Now suppose $i \neq j$, $k \neq l$ are integers and let $P \in \mathrm{GL}(\boldsymbol{Z})$ be the permutation matrix corresponding to the permutation $(ik)(jl)$. Let Q be the image of P in $\mathrm{GL}(A) \subset \mathrm{GL}(B)$ under the unique map $\boldsymbol{Z} \to A$ of rings with unit. Then Q is also a

permutation matrix and $Qe_{ij}(b)Q^{-1} = e_{kl}(b)$, $Qe_{ji}(b)Q^{-1} = e_{lk}(b)$ for all $b \in B$. It follows that $Qe_{ij}(b_1, ..., b_n; x)Q^{-1} = e_{kl}(b_1, ..., b_n; x)$ so

$$e_{kl}(b_1, ..., b_n; x)\, e_{ij}(b_1, ..., b_n; x)^{-1} = [Q, e_{ij}(b_1, ..., b_n; x)] \in [\mathrm{GL}(A), \mathrm{GL}(A, I)]$$
$$= E(A, I)$$

by [14, Th. 15.1]. Therefore $e_{ij}(b_1, ..., b_n; x)$ and $e_{kl}(b_1, ..., b_n; x)$ have the same image in $K_1(A, I)$. I will denote this image by $\epsilon(b_1, ..., b_n; x)$ since it is independent of the choice of i, j.

Corollary 4.4. *Let A be a subring of B, both rings having the same unit. Let $I \in A$ be an ideal of B. Then the kernel of the epimorphism $K_1(A, I) \to K_1(B, I)$ is generated by the elements $\epsilon(b_1, ..., b_n; x)$ for all $n \geq 0$, $b_1, ..., b_n \in B$, and $x \in I$.*

A less explicit form of this result says that the kernel of $K_1(A, I) \to K_1(B, I)$ is generated by the images in $K_1(A, I)$ of all $Ee_{12}(x)E^{-1}$ for $x \in I$, $E \in E_2(B)$. We can even allow all $E \in \mathrm{GL}_2(B)$. The $\epsilon(b_1, ..., b_n; x)$ all have this form and all these elements lie in the kernel since $e_{12}(x)$ has image 0 in $K_1(A, I)$ and $[E, e_{12}(x)] \in [\mathrm{GL}(B), \mathrm{GL}(B, I)] = E(B, I)$. It follows that there is an exact sequence

$$E_2(B, I) \to K_1(A, I) \to K_1(B, I) \to 0.$$

This may be interpreted as a stability theorem for the kernel of an excision map.

The kernel of $K_1(A, I) \to K_1(B, I)$ is also generated by the images in $K_1(A, I)$ of the elements $[e_{12}(b), x]$ for all $b \in B$, $x \in E_2(B, I)$. We can even allow all $x \in \mathrm{GL}_2(B, I) = \mathrm{GL}_2(A, I)$. Clearly all these elements lie in $E_2(B, I)$. Modulo these elements we have $\epsilon(b_1, ..., b_n; x) \equiv \epsilon(b_2, ..., b_n; x) \equiv ... \equiv \epsilon(x) = 0$.

We now consider the case where B is commutative. For any $r, s \in B$, the matrix $N(r, s) = \begin{pmatrix} rs & r^2 \\ -s^2 & -rs \end{pmatrix}$ satisfies $N(r, s)^2 = 0$. Therefore $U(r, s; x) = 1 + xN(r, s)$ is unipotent and lies in $\mathrm{GL}_2(B)$ for any $x \in B$. An easy calculation shows that

$$e_{12}(b)\, U(r, s; x)\, e_{12}(b)^{-1} = U(r - bs, s, x),$$

$$e_{21}(b)\, U(r, s; x)\, e_{21}(b)^{-1} = U(r, s - br, x).$$

It follows that the elements $e_{12}(b_1, ..., b_n; x)$ all have the form $U(r, s; x)$ for $x \in I$. Since

$$U(r, s; x) = \begin{pmatrix} 1+rsx & r^2x \\ -s^2x & 1-rsx \end{pmatrix}$$

lies in $\mathrm{SL}_2(A, I)$ for $x \in I$, we may take its image $u(r, s; x)$ in $\mathrm{SK}_1(A, I)$ and we have

$$u(r,s;x) = \begin{bmatrix} r^2x \\ 1+rsx \end{bmatrix}.$$

In particular, we see that $\epsilon(b,x) = u(1,-b;x) = [{}_{1-bx}^{\ x}]$. Now

$$u(r,s;x)\,\epsilon(-rs,x) = \begin{bmatrix} r^2x \\ 1+rsx \end{bmatrix}\begin{bmatrix} x \\ 1+rsx \end{bmatrix} = \begin{bmatrix} r^2x^2 \\ 1+rsx \end{bmatrix} = \begin{bmatrix} rx \\ 1+rsx \end{bmatrix}^2$$

$$= \epsilon(-rs, rx)^2.$$

Therefore $u(r,s;x) = \epsilon(-rs,x)^{-1}\,\epsilon(-rs,rx)^2$ so the subgroup of $K_1(A,I)$ generated by all $u(r,s;x)$ is the same as the subgroup generated by all $\epsilon(b,x)$. Since all $\epsilon(b_1,\dots,b_n;x)$ have the form $u(r,s;x)$, this subgroup is the kernel of $K_1(A,I) \to K_1(B,I)$. Further simplifications result from the following lemma.

Lemma 4.5. *For $b \in B$, $x \in I$, $\epsilon(b,x) = [{}_{1-bx}^{\ x}]$ is additive in b and in x. If $b \in A$, then $\epsilon(b,x) = 0$. If $x, y \in I$, then $\epsilon(b,xy) = 0$.*

Proof. By (S1) we see that $e_{ij}(b_1,\dots,b_n;x+y) = e_{ij}(b_1,\dots,b_n;x)\cdot e_{ij}(b_1,\dots,b_n;y)$. Therefore $\epsilon(b,x)$ is additive in x. Now

$$\epsilon(b_1,x)\,\epsilon(b_2,x) = \begin{bmatrix} x \\ 1-b_1x \end{bmatrix}\begin{bmatrix} x \\ 1-b_2x \end{bmatrix}$$

$$= \begin{bmatrix} x \\ (1-b_1x)(1-b_2x) \end{bmatrix} = \begin{bmatrix} x \\ 1-b_1x-b_2x+b_1b_2x^2 \end{bmatrix}.$$

Using (MS1b), we can subtract $(b_1b_2x)x$ from the denominator since $b_1b_2x \in I \subset A$. Therefore we get

$$\epsilon(b_1,x)\,\epsilon(b_2,x) = \begin{bmatrix} x \\ 1-b_1x-b_2x \end{bmatrix} = \epsilon(b_1+b_2,x).$$

If $b \in A$, then $[{}_{1-bx}^{\ x}] = 1$ by (MS1b). Finally

$$\epsilon(b,xy) = \begin{bmatrix} xy \\ 1-bxy \end{bmatrix} = \begin{bmatrix} x \\ 1-bxy \end{bmatrix}\begin{bmatrix} y \\ 1-bxy \end{bmatrix} = 1$$

by (MS1b) since $by, bx \in I \subset A$.

The following theorem summarizes the above results.

Theorem 4.6. *Let B be a commutative ring and let A be a subring of B. Let I be an ideal of B which lies in A. Then there is an exact sequence*

$$(B/A)\otimes(I/I^2) \xrightarrow{\psi} K_1(A,I) \to K_1(B,I) \to 0.$$

The tensor product here is over $\mathbf{Z}$ and ψ is induced by the map $B \otimes I \to K_1(A, I)$ sending $b \otimes x$ to $\epsilon(b, x) = [{}_1{}^{x}{}_{-bx}]$.

There is no need to assume that A or B has a unit since [17] $K_1(A, I) = K_1(A^+, I)$, $K_1(B, I) = K_1(B^+, I)$ and $B^+/A^+ = B/A$.

Corollary 4.7. *Let $f: A \to B$ be a homomorphism of commutative rings. Let $I \subset A$, $J \subset B$ be ideals such that $f: I \approx J$. Then there is an exact sequence*

$$(B/f(A)) \otimes I/I^2 \to K_1(A, I) \to K_1(B, J) \to 0.$$

This sequence is natural with respect to (A, B, I, J, f).

We need only observe that $K_1(A, I) = K_1(f(A), J)$. The naturality will be exploited in §5.

Theorem 4.6 gives us a simple universal example for excision. If R is a commutative ring with unit, let $\Gamma_R = R[u, v]$, $\mathfrak{A}_R = \Gamma_R v$, and $\Lambda_R = R + \mathfrak{A}_R$. Let $\epsilon_R = \epsilon(u, v) \in K_1(\Lambda_R, \mathfrak{A}_R)$. This lies in the kernel of $K_1(\Lambda_R, \mathfrak{A}_R) \to K_1(\Gamma_R, \mathfrak{A}_R)$. In the situation of Theorem 4.6, if A and B are R-algebras each $\epsilon(b, x)$ with $b \in B$, $x \in I$ is the image of ϵ_R under the map $\Gamma_R \to B$ sending u to b and v to x. Thus if $\epsilon_R = 0$, excision will hold in the case of commutative R-algebras. It is quite reasonable to conjecture that $\epsilon_R \neq 0$ for every $R \neq 0$. To prove this, it will be sufficient to consider the case where R is a finite field. This is shown by the following argument.

Write $\Lambda = \Lambda_{\mathbf{Z}}$, $\Gamma = \Gamma_{\mathbf{Z}}$, $\mathfrak{A} = \mathfrak{A}_{\mathbf{Z}}$, and $\epsilon = \epsilon_{\mathbf{Z}}$. Then $\Lambda_R = R \otimes \Lambda$, $\Gamma_R = R \otimes \Gamma$, $\mathfrak{A}_R = R \otimes \mathfrak{A}$ and ϵ_R is the image of ϵ under the canonical map $\Gamma \to \Gamma_R$. Since K_1 preserves filtered direct limits, we can express R as a filtered union of finitely generated subrings R_α and $K_1(R \otimes \Lambda, R \otimes \mathfrak{A}) = \varinjlim K(R_\alpha \otimes \Lambda, R_\alpha \otimes \mathfrak{A})$. If $\epsilon_R = 0$, the image of ϵ in this limit is zero and it follows that $\epsilon_{R_\alpha} = 0$ for some α. Therefore we can assume that R is a finitely generated ring. Let $\mathfrak{M}$ be a maximal ideal of R. Then ϵ_R maps to $\epsilon_{R/\mathfrak{M}}$ under the map induced by $R \to R/\mathfrak{M}$. We now apply the following result which is presumably well-known.

Lemma 4.8. *A field which is finitely generated as a ring is finite.*

Proof. Let F be the field and let $k \subset F$ be the prime field. Hilbert's nullstellensatz implies that F is a finite algebraic extension of k [19, p. 165] so we are done unless $k = \mathbf{Q}$. In this case, F is an algebraic number field. If so, let R be the ring of integers of F. Then $\mathbf{Q}R = F$. If F is generated as a ring by $x_1, \ldots, x_n$, write $x_i = a_i/m_i$ where $a_i \in R$, $m_i \in \mathbf{Z}$. Then $F = \mathbf{Z}[x_1, \ldots, x_n] = R[N^{-1}]$ where $N = \prod m_i$. If p is a prime not dividing N and $\mathfrak{p}/p$ is a prime of R, it follows that $\operatorname{ord}(x) \geq 0$ for all $x \in F$. This is impossible for $x = p^{-1}$.

We can also get universal examples for excision in the non-commutative case. Let $R_n = R\{y_1, \ldots, y_n; x\}$ be the non-commutative polynomial ring on the indicated

indeterminates. Let I_n be the 2-sided ideal of R_n generated by x and let $R'_n = R + I_n$. We consider $\epsilon(y_1, ..., y_n; x)$ in $K_1(R'_n, I_n)$. These elements are all non-zero by §1.

§5. Rings of arithmetic type

I will now show how Theorem 4.6 can be applied to extend the qualitative results of [2] to rings of "arithmetic type" which are not integrally closed. It would be very interesting to have a corresponding extension of the quantitative results of [2] which give the precise order of $SK_1(R, I)$.

Let R be an integral domain with quotient field K. We assume that K is either algebraic over Q or that K has characteristic $p \neq 0$ and has transcendence degree $\leqslant 1$ over the prime field. Let μ_K denote the group of roots of unity of K. Let $I \subset R$ be any ideal. We do not assume K finitely generated over the prime field.

Theorem 5.1. *Under these conditions,* $SK_1(R, I)$ *is isomorphic to a subgroup of* μ_K, *and* $SK_1(R, I) = 0$ *except in the following case*

(A) *K is a totally imaginary field of characteristic zero and R is integral over Z. In case* (A) *if R is finite over Z, there is a non-zero ideal J of R such that* $SK_1(R, I) = \mu_K$ *whenever* $0 \neq I \subset J$.

It is not necessary to assume that $1 \in R$ here since $SK_1(R, I) = SK_1(R^+, I) = SK_1(R', I)$ where R' is the image of R^+ in K, $R' = Z \cdot 1 + R$. For the proof, I will, of course, assume that R has a unit. If K has finite characteristic and is algebraic over the prime field k then $R = K$ and the theorem is trivial. Therefore we can ignore this case. We first reduce to the case where R is finitely generated. Express R as a filtered union of finitely generated subrings R_α and let $I_\alpha = I \cap R_\alpha$. Then $SK_1(R, I) = \varinjlim SK_1(R_\alpha, I_\alpha)$.

We note first that if case (A) does not apply to R, the same is true for all sufficiently large R_α. Clearly the characteristics are the same. If $r \in R$ is not integral over Z, then R_α is not integral over Z as soon as $r \in R_\alpha$. If K is not totally imaginary, it has a real conjugate and the same is true of all R_α. If we can prove the theorem for the case where R is finitely generated, it follows immediately that $SK_1(R, I) = 0$ except in case (A). To see that $SK_1(R, I)$ is isomorphic to a subgroup of μ_K we use the following result.

Lemma 5.2. *If* $G = \varinjlim G_\alpha$ *is a filtered direct limit of Abelian groups and each* G_α *is isomorphic to a subgroup of* μ_K *then the same is true for* G.

Proof. Since μ_K is a torsion group, the same is true of G_α and G and we can consider the p-primary components separately. Assume all G_α are p-groups. The p-component of μ_K is either Z/p^nZ for some n or the group Z_{p^∞} which is the only infinite Abelian group which is a filtered union of cyclic p-groups. If $x, y \in G$ then x and y are both

in the image of G_α for some α. It follows that any finitely generated subgroup of G is a cyclic p-group and so G is either $\mathbf{Z}_{p^\infty}$ or $\mathbf{Z}/p^m\mathbf{Z}$ for some m. If $\mu_K \approx \mathbf{Z}/p^n\mathbf{Z}$, all elements of G_α have order dividing p^n. The same is then true of G so $m \leqslant n$ and G is isomorphic to a subgroup of $\mathbf{Z}/p^n\mathbf{Z}$.

We can now assume that R is a finitely generated ring. It follows that K is a global field.

Lemma 5.3. *Let R be a domain whose quotient field K is a global field. Let A be the integral closure of R in K. Then A is finite over R. If R is a finitely generated ring, then A is a ring of arithmetic type in the sense of [2].*

Proof. If K has characteristic zero, let B be the ring of integers of K. Since B is finite over $\mathbf{Z}$, the ring RB is finite over R. Since $B \subset RB$, it follows from [15, A22] that $RB = B_S$ for some multiplicative set S in B. Thus RB is a Dedekind ring and so is integrally closed. It follows that RB is the integral closure of R. If K has finite characteristic, the proof of [15, A23] shows that we can find $x \in R$ such that K is separable over $k(x)$ where k is the prime field. Now $k[x] \subset R$ and the integral closure B of $k[x]$ in K is finite over $k[x]$ [15, A9]. As above, it follows that RB is the integral closure of R and is finite over R.

If R is finitely generated, so is its integral closure A which is finite over R. Let A be generated by $a_1, ..., a_n$. If $\mathfrak{p}$ is a prime of K such that $\mathrm{ord}_\mathfrak{p}\, a_i \geqslant 0$ for all i, then $A \subset \mathcal{O}_\mathfrak{p}$ and conversely. Since A is Dedekind, $A = \cap \mathcal{O}_\mathfrak{p}$ over such $\mathfrak{p}$ and all but a finite number of primes of K occur.

We can now prove Theorem 5.1. Since $K_1(R, 0) = 0$ we look only at non-zero ideals I. If $0 \neq I \subset J$ are ideals of R, then by [1, VI, Prop. 1.4], $\mathrm{SK}_1(R, I) \to \mathrm{SK}_1(R, J)$ is onto. Clearly R is a Noetherian domain and $\dim R = \dim A = 1$. Therefore, we need only look at sufficiently small ideals I. Note that any quotient of a subgroup of μ_K is again isomorphic to a subgroup of μ_K since its p-primary components are cyclic.

Since A is finite over R we can find $r \neq 0$ in R such that $rA \subset R$. If I is an ideal of R, then $0 \neq rAI \subset I$ and rAI is an ideal of A. Since $\mathrm{SK}_1(R, rAI) \to \mathrm{SK}_1(R, I)$ is onto, we need only look at ideals of R which are also ideals of A.

If case (A) applies, then by [2, Th. 3.6] there is an ideal J_0 of A such that $\mathrm{SK}_1(A, J) = \mu_K$ for all ideals J of A with $J \subset J_0$. If $J_2 \subset J_1 \subset J_0$, then $\mathrm{SK}_1(A, J_2) \to \mathrm{SK}_1(A, J_1)$ is onto and therefore an isomorphism. If case (A) does not apply we set $J_0 = A$, since all $\mathrm{SK}_1(A, J)$ are zero by [2, Th. 3.6].

Let $rA \subset R$ with $r \neq 0$ as above and let $J = (rJ_0)^2$. Let $I \neq 0$ be an ideal of A contained in R with $I \subset J$ and consider the following diagram obtained from Theorem 4.6 by using the fact that all $\epsilon(r, x)$ lie in SK_1,

$$\begin{array}{ccccccc}
A/R \otimes I^2/I^4 & \longrightarrow & \mathrm{SK}_1(R, I^2) & \longrightarrow & \mathrm{SK}_1(A, I^2) & \longrightarrow & 0 \\
\downarrow{\scriptstyle 0} & & \downarrow & & \downarrow{\scriptstyle \approx} & & \\
A/R \otimes I/I^2 & \longrightarrow & \mathrm{SK}_1(R, I) & \longrightarrow & \mathrm{SK}_1(A, I) & \longrightarrow & 0\,.
\end{array}$$

The left vertical map is obviously zero since it is induced by the inclusion $I^2 \subset I$. The right vertical map is an isomorphism since $I^2 \subset I \subset J_0$. The middle vertical map is onto. An easy diagram chase now shows that $SK_1(R, I) \to SK_1(A, I)$ is an isomorphism. Since $SK_1(A, I) = \mu_K$ in case (A) and $SK_1(A, I) = 0$ otherwise, we are done.

Remark. In the hope of inspiring someone to extend the quantitative results of [2], I should point out that Theorem 5.1 can also be obtained using the methods of [2]. Using the above remarks and [1, VI, Th. 2.3] we are reduced to a rather tedious reworking of the methods of [2, Part I]. There is only one point worth mentioning explicitly. In order to find a prime element π of R by Dirichlet's theorem we look instead for a prime element of A which satisfies the required conditions and, in addition, the condition $\pi \equiv 1 \bmod I$ where $I \subset R$ is an ideal of A as above. Now if $a\pi \in R$ with $a \in A$, we have $a \equiv a\pi \bmod I$ and so $a \in R + I = R$. In particular $A\pi \cap R = R\pi$ so $R\pi$ is prime in R.

I have not attempted to prove anything beyond the results of Theorem 5.1 by this method but it is quite likely that more precise results could be obtained at least if $I \subset R$ is an ideal of A. The case where I is not an ideal of A will probably be even more difficult. In particular, I do not know whether $SK_1(R) = 0$ in case (A).

§6. Failure of excision for K_2

A number of definitions have been proposed for $K_n(R)$, $n \geqslant 2$ [5, 9, 10, 16]. I will show here that in all of these theories excision fails completely for K_2. Rather than consider each separately, I will state a number of axioms which are satisfied by all these theories and deduce the results from these axioms. The results do not apply to the K-theory of Karoubi and Villamayor [6, 8] for which excision is known to hold. The axioms below are not satisfied by this theory because its K_1 is not the usual one.

I will denote Milnor's K_2 by MK_2 here to distinguish it from the other K_2 under consideration. Thus $MK_2(R) = \ker [\mathrm{St}(R) \to \mathrm{GL}(R)]$. Suppose we are given a functor K_2 from rings with unit to Abelian groups and also a relative functor K_2 defined on the category of pairs (R, I) where R is a ring with unit and I is a 2-sided ideal of R. The values of K_2 are to be in the category of Abelian groups. We assume that the following three conditions are satisfied.

(A) There is a natural map θ: $MK_2 \to K_2$.

(B) If I is a 2-sided ideal of R, there is a map ∂: $K_2(R/I) \to K_1(R, I)$ such that the diagram

$$\begin{tikzcd} MK_2(R/I) \arrow[r, "\partial"] \arrow[d, "\theta"'] & K_1(R, I) \arrow[d, equal] \\ K_2(R/I) \arrow[r, "\partial"] & K_1(R, I) \end{tikzcd}$$

commutes. Here $K_1(R, I)$ is the usual relative K_1 and the map ∂: $\mathrm{MK}_2(R/I) \to K_1(R, I)$ is the usual map in the exact sequence

$$(1) \qquad \mathrm{MK}_2(R, I) \to \mathrm{MK}_2(R) \to \mathrm{MK}_2(R/I) \xrightarrow{j} K_1(R, I) \to K_1(R) \to K_1(R/I) \to \dots .$$

(C) There is a natural map ι: $K_2(R, I) \to K_2(R)$ such that ι is a monomorphism whenever $R \to R/I$ is a split epimorphism of rings.

In the theories of [5, 10, 16] there is actually a sequence

$$(2) \qquad \dots \to K_3(R) \to K_3(R/I) \to K_2(R, I) \to K_2(R) \to K_2(R/I) \to K_1(R, I) \to \dots$$

and a natural map of (1) into this which is the identity on the K_1 terms. If we take $K_2 = \mathrm{MK}_2$, we can define $K_2(R, I)$ to be the kernel of $K_2(R) \to K_2(R/I)$. This was denoted $K'_2(R, I)$ in [14]. In the other theories (C) follows immediately from (2).

In [9], Milnor defines $\mathrm{MK}_2(R, I)$ to be the Stein relativation of MK_2. We will see that this does not satisfy (C) and so does not fit into an exact sequence (2). For any K_2 as above, we define $K_2^{\$}(R, I)$ to be the Stein relativation of K_2. This is defined as follows [12]. Let

$$\begin{tikzcd} D \arrow[r, "p_1"] \arrow[d, "p_2"'] & R \arrow[d] \\ R \arrow[r] & R/I \end{tikzcd}$$

be a Cartesian diagram, and define $K_2^{\$}(R, I) = \ker [K_2(D) \to K_2(R)]$. By (C) we have $K_2^{\$}(R, I) = K_2(D, J)$ where J is the kernel of p_1 because p_1 (as well as p_2) is split by the diagonal map Δ: $R \to D$. The map p_2: $(D, J) \to (R, I)$ induces a map $K_2^{\$}(R, I) = K_2(D, J) \to K_2(R, I)$. I will show that this is not an isomorphism in general and that $K_2^{\$}(R, I)$ does not satisfy (C). I will also show that excision fails for $K_2^{\$}(R, I)$.

The obstructions to excision are given by a certain secondary K-theory operation. We begin by defining this and showing that it is non-trivial.

Let R be a ring with unit. If $a, b \in R$ satisfy $ab = 0 = ba$, then in $\mathrm{GL}(R)$ we have $e_{12}(a)\, e_{21}(b) = 1 + ae_{12} + be_{21} = e_{21}(b)\, e_{12}(a)$. Therefore the element $[x_{12}(a), x_{21}(b)]$ lies in the kernel $\mathrm{MK}_2(R)$ of φ: $\mathrm{St}(R) \to \mathrm{GL}(R)$. In the notation of [9, §8] it is $e_{12}(a) * e_{21}(b)$.

Definition. If $a, b \in R$ and $ab = 0 = ba$, let $c(a, b) = \theta([x_{12}(a), x_{21}(b)]) \in K_2(R)$.

This is clearly natural. If f: $R \to R'$ then $K_2(f)(c(a, b)) = c(f(a), f(b))$. We also note that $c(a, b) = 0$ if a or b is 0. The ring $\mathbf{Z}[x, y]/(xy)$ is clearly a universal example for the operation c: If $ab = ba = 0$ in R, define f: $\mathbf{Z}[x, y]/(xy) \to R$ by $f(x) = a$, $f(y) = b$. Then $c(a, b) = K_2(f)(c(x, y))$.

Theorem 6.1. *$c(x, y) \neq 0$ in $K_2(\mathbf{Z}[x, y]/(xy))$.*

It would be interesting to know whether this result extends to the case where $\mathbf{Z}$ is replaced by some other commutative ring R, e.g., a field. This would imply that all the results obtained below continue to hold if we restrict ourselves to the category of R-algebras. At the end of this section I will show how to do this for the fields $\mathbf{Z}/p\mathbf{Z}$ with $p \geqslant 5$.

Since $c(x, y)$ is universal, it will clearly suffice to find some ring R for which $c(a, b) \neq 0$. We begin by reducing our problem to one concerning K_1.

Lemma 6.2. *Let R be a commutative ring with unit and let I be an ideal in R. Let $a, b \in R$ with $ab \in I$. If $\bar{a}, \bar{b}$ are the images of a, b in R/I, then the image of $c(\bar{a}, \bar{b})$ under the map ∂ given in* (B) *is the Mennicke symbol $\left[{ab^2 \atop 1-ab}\right] \in \mathrm{SK}_1(R, I)$.*

Proof. The commutative diagram of (B) shows that $\partial c(\bar{a}, \bar{b})$ is the image of $[x_{12}(\bar{a}), x_{21}(\bar{b})]$ under the map $\partial\colon \mathrm{MK}_2(R/I) \to K_1(R, I)$ in sequence (1). This sequence is obtained by applying the snake lemma to the diagram

$$\begin{array}{ccccccc} 1 & \to & \mathrm{St}'(R, I) & \to & \mathrm{St}(R) & \to & \mathrm{St}(R/I) & \to & 1 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 1 & \to & \mathrm{GL}(R, I) & \to & \mathrm{GL}(R) & \to & \mathrm{GL}(R/I) & . & \end{array}$$

Lift $[x_{12}(\bar{a}), x_{21}(\bar{b})]$ back to $[x_{12}(a), x_{21}(b)]$ in $\mathrm{St}(R)$. The image of this in $\mathrm{GL}(R)$ is

$$[e_{12}(a), e_{21}(b)] = \begin{pmatrix} 1+ab+a^2b^2 & a^2b \\ ab^2 & 1-ab \end{pmatrix}.$$

This lies in $\mathrm{GL}(R, I)$ and represents $\partial c(\bar{a}, \bar{b})$. Since $[e_{12}(a), e_{21}(b)]$ is a commutator, it lies in $\mathrm{SL}(R, I)$. We now apply the following well-known result (cf. [1, p. 300]).

Lemma 6.3. *The image in $K_1(R, I)$ of $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(R, I)$ is $\left[{b \atop a}\right] = \left[{b \atop d}\right]^{-1} = \left[{c \atop d}\right] = \left[{c \atop a}\right]^{-1}$.*

Proof. The image is $\left[{b \atop a}\right]$ by definition. Since $ad - bc = 1$, we have $\left[{b \atop a}\right]\left[{b \atop d}\right] = \left[{b \atop ad}\right] = \left[{b \atop 1+bc}\right] = \left[{b \atop 1}\right] = \left[{0 \atop 1}\right] = 1$ and similarly $\left[{c \atop d}\right] = \left[{c \atop a}\right]^{-1}$. Also $\left[{b \atop a}\right]\left[{c \atop a}\right] = \left[{bc \atop a}\right] = \left[{ad-1 \atop a}\right] = \left[{a(d-1)+a-1 \atop a}\right] = \left[{a-1 \atop a}\right] = \left[{a-1 \atop 1}\right] = 1$.

Now let ζ be a primitive cube root of 1 and let $A = \mathbf{Z}[\zeta]$. Let $\lambda = 1 - \zeta$. Then the ideal (λ) of A is prime and $(\lambda)^2 = (3)$. In fact, $\lambda^2 = -3\zeta$. Let $I = (\lambda^3)$ and set $a = \lambda^2$, $b = -\zeta^2\lambda$. Then $ab \in I$.

Lemma 6.4. *For this particular A we have $\left[{ab^2 \atop 1-ab}\right] \neq 1$ in $K_1(A, I)$ and so $c(\bar{a}, \bar{b}) \neq 0$ in $K_2(A/I)$.*

Proof. By [2, Th. 3.6], there is a homomorphism $\mathrm{SK}_1(A, I) \to \mu_3$ given by $\left[\begin{smallmatrix}\beta\\ \alpha\end{smallmatrix}\right] \mapsto \left(\frac{\beta}{\alpha}\right)_3$. Now $ab^2 = \zeta\lambda^4 = 9$ and $1 - ab = 1 + \zeta^2\lambda^3 = 1 - 3\lambda = -2 + 3\zeta$. Since $N(1 - 3\lambda) = 19$, we see that $(1 - 3\lambda)$ is prime and $A/(1 - 3\lambda) = \boldsymbol{F}_{19}$. Since $(19 - 1)/3 = 6$, we have by definition, $\left(\frac{9}{1-3\lambda}\right) \equiv 9^6 \equiv -8 \bmod (1 + 3\lambda)$. Therefore this element is non-trivial.

We now apply Theorem 6.1 to prove a number of negative results concerning K_2. Consider the Cartesian diagram

$$
\text{(3)} \qquad
\begin{tikzcd}
\boldsymbol{Z}[x,y]/(xy) \arrow[r, "p_1"] \arrow[d, "p_2"'] & \boldsymbol{Z}[t] \arrow[d, "f"] \\
\boldsymbol{Z}[t] \arrow[r, "f"'] & \boldsymbol{Z}
\end{tikzcd}
$$

where $f(t) = 0$, $p_1(x) = t$, $p_1(y) = 0$, $p_2(x) = 0$, $p_2(y) = t$. For convenience, we write $R = \boldsymbol{Z}[t]$, $I = \ker f$, $D = \boldsymbol{Z}[x,y]/(xy)$, and $J = \ker p_1$. This notation is used in all the following corollaries.

Corollary 6.5. *The map $K_2^{\mathrm{S}}(R, I) \to K_2(R, I)$ is not an isomorphism.*

Proof. By definition $K_2^{\mathrm{S}}(R, I) = K_2(D, J) = \ker [K_2(p_1): K_2(D) \to K_2(R)]$. The image of $c(x, y)$ under $K_2(p_1)$ is $c(t, 0) = 0$. Thus $c(x, y) \in K_2(D, J)$. Its image under $K_2(p_2)$ is $c(0, t) = 0$ so $c(x, y)$ lies in the kernel of $K_2(D, J) \to K_2(R, I)$.

It follows that the sequence $0 \to K_2^{\mathrm{S}}(R, I) \to K_2(R) \to K_2(R/I) \to 0$ is not exact even if $R \to R/I$ is a split epimorphism. In particular there is no sequence of the form (2) with $K_2(R, I) = K_2^{\mathrm{S}}(R, I)$.

Corollary 6.6. *The map $K_2(D, J) \to K_2(R, I)$ is not an isomorphism even though $p_2: (D, J) \to (R, I)$ is a split epimorphism and $p_2: J \approx I$.*

Thus excision fails for K_2. We have just shown that $K_2(D, J) \to K_2(R, I)$ is not an isomorphism. Clearly $p_2: J \approx I$. Define $g: \boldsymbol{Z}[t] \to \boldsymbol{Z}[x, y]/(xy)$ by $g(t) = y$. Then $p_2 g$ is the identity. Also $g(I) \subset J$ since $I = Rt$ and $J = Dy$. Therefore $(D, J) \to (R, I)$ is a split epimorphism.

Corollary 6.7. *The map $K_2^{\mathrm{S}}(D, J) \to K_2^{\mathrm{S}}(R, I)$ is not an isomorphism.*

Therefore excision also fails for $K_2^{\mathrm{S}}(R, I)$ even under split epimorphisms.

Proof. To find $K_2^{\mathrm{S}}(D, J)$ we must consider the Cartesian diagram

$$
\text{(4)} \qquad
\begin{tikzcd}
E \arrow[r, "q_1"] \arrow[d, "q_2"'] & D \arrow[d, "p_1"] \\
D \arrow[r, "p_1"'] & D/J
\end{tikzcd}
$$

where we have identified $D \to D/J$ with $p_1 : D \to R$. If we regard (3) as a map of $p_1 : D \to R$ to $f: R \to Z$, and apply Stein's construction, we get a map of (4) into (3). Explicitly, this is

(5)
$$\begin{tikzcd}
& D \arrow[rr, "f"] \arrow[dd, "p_2"] & & R \arrow[dd, "f"] \\
E \arrow[ur, "h"] \arrow[rr, "q_1"] \arrow[dd, "q_2"] & & D \arrow[ur, "p_2"] \arrow[dd, "p_1"] & \\
& R \arrow[rr, "f"] & & Z \\
D \arrow[ur, dashed, "p_2"] \arrow[rr, "p_1"] & & D/J \arrow[ur, "f"] &
\end{tikzcd}$$

where $h(d_1, d_2) = (p_2(d_1), p_2(d_2))$. Since $K_2^s(R, I) = \ker K_2(p_1)$ and $K_2^s(D, J) = \ker K_2(q_1)$, the top square of (5) gives us the diagram

$$\begin{tikzcd}
0 \arrow[r] & K_2^s(D,J) \arrow[r] \arrow[d] & K_2(E) \arrow[r, "q_1"] \arrow[d, "h"] & K_2(D) \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & K_2^s(R,I) \arrow[r] & K_2(D) \arrow[r, "p_1"] & K_2(R) \arrow[r] & 0 .
\end{tikzcd}$$

Consider the elements $a = (x, x)$, $b = (0, y)$ in E. Then $ab = 0$ so $c(a, b) \in K_2(E)$ is defined. Since $q_1(b) = 0$ and $h(a) = 0$ we see that $c(a, b)$ lies in the kernel of $K_2^s(D, J) \to K_2^s(R, I)$ just as in Corollary 6.5. To show that $c(a, b) \neq 0$ we observe that $q_2(a) = x$ and $q_2(b) = y$. Therefore the image of $c(a, b)$ under $K_2(q_2)$ is $c(x, y) \neq 0$.

If we have a map $f: (R_1, I_1) \to (R_2, I_2)$ with $f: I_1 \approx I_2$ and if $f: R_1 \to R_2$ is onto, it is easy to see that the map $\mathrm{MK}_2(R_1, I_1) \to \mathrm{MK}_2(R_2, I_2)$ is onto [9, Lemma 6.3]. The condition that f be onto is really needed here. Let g be as in the proof of Corollary 6.6.

Corollary 6.8. *$g: (R, I) \to (D, J)$ has $g: I \approx J$ but $K_2(R, I) \to K_2(D, J)$ and $K_2^s(R, I) \to K_2^s(D, J)$ are not onto.*

In fact, $K_2(p_2)\, K_2(g) = \mathrm{id}$. If $K_2(g)$ were onto, it would follow that $K_2(p_2)$ is an isomorphism.

We can also get a negative result in the absolute case. Here we only need to assume that K_2 satisfies (A) and (B).

Corollary 6.9. *There is no functor K_3 from rings to Abelian groups such that for every Cartesian diagram*

$$\begin{tikzcd} A \arrow[r] \arrow[d] & A_1 \arrow[d, "g"] \\ A_2 \arrow[r, "f"'] & A' \end{tikzcd}$$

with f and g split epimorphisms, there is an exact Mayer-Vietoris sequence

$$K_3(A_1) \oplus K_3(A_2) \to K_3(A') \to K_2(A) \to K_2(A_1) \oplus K_2(A_2).$$

In fact, since f and g are split, the existence of such a sequence would imply that $K_2(A) \to K_2(A_1) \oplus K_2(A_2)$ is a monomorphism. But in the case of diagram (3), the element $c(x, y)$ of $K_2(D)$ is non-zero and maps to zero under p_1 and p_2.

Finally, we consider the extension of K_2 to rings without units considered in [16, 17]. If A has no unit, let A^+ be the ring obtained by formally adjoining a unit to A. We have a split extension $0 \to A \to A^+ \to \mathbf{Z} \to 0$ and we define $K_2(A) = \ker [K_2(A^+) \to K_2(\mathbf{Z})]$. This is consistent if A has a unit provided that K_2 preserves finite products (when applied to rings with units). This is certainly true for MK_2.

Corollary 6.10. *The extension of K_2 to rings without unit does not preserve finite products.*

Proof. Using the diagram (3) again, we see that $R = I^+$ and $D = (I \times I)^+$. Since x and y map to zero in $\mathbf{Z}$, $c(x, y)$ lies in $K_2(I \times I)$. Its images under p_1 and p_2 are zero so $K_2(I \times I) \to K_2(I) \times K_2(I)$ is not a monomorphism. Note that $K_2(R_1 \times R_2) \to K_2(R_1) \times K_2(R_2)$ is always a split epimorphism because the inclusions i_ν: $R_\nu \to R_1 \times R_2$ given by $i_1(r) = (r, 0)$, $i_2(r) = (0, r)$ are ring homomorphisms. Since $K_2(0) = 0$ it is easy to check that the composition $K_2(R_1) \times K_2(R_2) \to K_2(R_1 \times R_2) \to K_2(R_1) \times K_2(R_2)$ is the identity.

To conclude this section, I will show that $c(x, y) \neq 0$ in $K_1(R(x, y)/(xy))$ for $A = \mathbf{Z}/p\mathbf{Z}$ where p is prime and $p \geqslant 5$. To do this it will suffice to exhibit a ring B and an ideal I with $a, b \in B$, $ab \in I$ and $[{}^{ab^2}_{1-ab}] \neq 1$ in $\mathrm{SK}_1(B, I)$ such that $p(B/I) = 0$. For this, we choose B and I as in Lemma 3.2. Let $a = \lambda^{p-3}xu^{-1}$, $b = \lambda^2 u$ where $\lambda = 1 - \zeta$, $x \in A$ and u is a unit of $A = \mathbf{Z}[\zeta]$. By [2, A11] there is an element $\pi \in A$ with $A\pi$ prime and $\pi \equiv 1 + p \pmod{p^2}$. Since $(p) = (\lambda^{p-1})$, we can choose $x \in A$ so that $1 - ab = 1 - \lambda^{p-1}x = \pi$. Now each conjugate $\pi^{(i)}$ of π satisfies $\pi^{(i)} \equiv 1 + p \pmod{p^2}$ so $N\pi = \prod \pi^{(i)} \equiv (1+p)^{p-1} \equiv 1 - p \pmod{p^2}$. Therefore $k = (N\pi - 1)/p \equiv -1 \pmod{p}$. Since π is prime,

$$\left(\frac{ab^2}{1-ab}\right)_p = \left(\frac{\lambda^{p+1}xu}{\pi}\right)_p \equiv (\lambda^{p+1}xu)^k \pmod{\pi}.$$

If $u = \zeta^r$, then $u^k = u^{-1}$ and we can choose r so that $(\lambda^{p+1}xu)^k \not\equiv 1 \bmod \pi$.

Using the remarks at the end of § 3, we can get many other examples and probably all finite fields of characteristic $p \neq 2, 3$ can be obtained in this way. The example considered in Lemma 7.4 shows that $c(x, y) \neq 0$ for $R = \mathbf{Z}/9\mathbf{Z}$ but I have not yet found an example for $R = \mathbf{Z}/3\mathbf{Z}$. An example for $R = \mathbf{Z}/4\mathbf{Z}$ may be obtained as follows. Let $A = \mathbf{Z}[\mathrm{i}]$, $B = \mathbf{Z} + 2A$, and $I = 2\lambda A$ where $\lambda = 1 - \mathrm{i}$, $\mathrm{i} = \sqrt{-1}$. Let $a = -2\mathrm{i}$ and $b = \lambda\mathrm{i}$. Again I have no example for $A = \mathbf{Z}/2\mathbf{Z}$.

§ 7. Further results on Milnor's K_2

In this section, I will give an analogue of Theorem 4.6 for Milnor's K_2. I will now denote this functor by K_2 rather than MK_2 as in §6. The Stein relativization of K_2 will be denoted by $K_2(R, I)$ as in [9] and $K_2'(R, I)$ will denote the kernel of $K_2(R) \to K_2(R/I)$.

Let $f: A \to B$ be a surjective homomorphism of rings with unit. Let $N = \ker f$ and let $I \subset A$ and $J \subset B$ be 2-sided ideals such that $f: I \approx J$. We can regard I and N as 2-sided A-modules and hence as modules over the ring $A^e = A \otimes A$.

Theorem 7.1. *If $f: A \to B$ is onto with kernel N and $f: I \approx J$, there is an exact sequence*

$$(N/N^2) \otimes_{A^e} (I/I^2) \xrightarrow{\psi} K_2(A, I) \to K_2(B, J) \to 0 .$$

The rings are allowed to be non-commutative here.

The map ψ is obtained as follows. Consider the Cartesian diagram

$$\begin{tikzcd} A_1 \arrow[r, "p_1"] \arrow[d, "p_2"'] & A \arrow[d] \\ A \arrow[r] & A/I \end{tikzcd}$$

If $x \in N$ and $y \in I$ then $xy, yx \in IN + NI \subset N \cap I = 0$. Therefore the elements (x, x) and $(0, y)$ of A_1 have product zero and we can define $\psi(x, y) = c((x, x), (0, y))$ in the notation of §6. I will show that ψ is A^e-bilinear and is zero if $x \in N^2$ or $y \in I^2$.

There is also an analogue of Theorem 7.1 for K_2' provided the extensions $A \to A/I$, $B \to B/J$ split in a consistent way.

Theorem 7.2. *Suppose $f: A \to B$ is onto, $N = \ker f$, and $f: I \approx J$. Suppose also that there are ring homomorphisms $\alpha: A/I \to A$, $\beta: B/J \to B$ which split the canonical maps $A \to A/I$, $B \to B/J$ such that the diagram*

$$\begin{tikzcd} A & A/I \arrow[l, "\alpha"'] \arrow[d, "f'"] \\ B \arrow[from=u, "f"'] & B/J \arrow[l, "\beta"] \end{tikzcd}$$

commutes, where f' is induced by f. Then the diagram

$$\begin{tikzcd} K_2(A,I) \arrow[r] \arrow[d, dashed] & K_2(B,J) \arrow[d] \\ K'_2(A,I) \arrow[r] & K'_2(B,J) \end{tikzcd}$$

is cocartesian.

It follows that the kernel of $K_2(A,I) \to K_2(B,J)$ maps onto the kernel of $K'_2(A,I) \to K'_2(B,J)$. Using Theorem 7.1 we deduce the following result.

Corollary 7.3. *Under the conditions of Theorem 7.2, there is an exact sequence*

$$(N/N^2)\otimes_{A^e}(I/I^2) \xrightarrow{\psi'} K'_2(A,I) \to K'_2(B,J) \to 0$$

where ψ' is induced by the map sending $x\otimes y$ to $c(x,y)$.

I will actually prove Corollary 7.3 and deduce Theorems 7.1 and 7.2 from it. To do this consider the Cartesian diagrams

$$\begin{tikzcd} A_1 \arrow[r, "p_1"] \arrow[d, dashed, "p_2"'] & A \arrow[d, dashed] \\ A \arrow[r] & A/I \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} B_1 \arrow[r, "q_1"] \arrow[d, "q_2"'] & B \arrow[d, dashed] \\ B \arrow[r] & B/I \end{tikzcd} .$$

Let $I_1 = \ker p_1$ and $J_1 = \ker q_1$ and let $f_1 : A_1 \to B_1$ by $f(a_1, a_2) = (f(a_1), f(a_2))$. If $f(a_1, a_2) = 0$ then $a_1, a_2 \in N$. Since $(a_1, a_2) \in A_1$ we have $a_1 - a_2 \in I$. But $I \cap N = 0$ so $a_1 = a_2$. Therefore $\ker f_1 = \{(a,a) | a \in N\} \approx N$. The diagonal maps $\Delta_A : A \to A_1$ and $\Delta_B : B \to B_1$ split the maps p_1 and q_1. All the hypotheses of Theorem 7.2 are satisfied by $f_1 : (A_1, I_1) \to (B_1, J_1)$. Therefore the exact sequence of Theorem 7.1 follows from that of Corollary 7.3. If in addition, the hypothesis of Theorem 7.2 applies to f, we get a diagram

$$\begin{tikzcd} (N/N^2)\otimes(I/I^2) \arrow[r, "\psi"] \arrow[d, dashed, "\mathrm{id}"] & K_2(A,I) \arrow[r] \arrow[d, dashed] & K_2(B,J) \arrow[r] \arrow[d, dashed] & 0 \\ (N/N^2)\otimes(I/I^2) \arrow[r, "\psi'"] & K'_2(A,I) \arrow[r] & K'_2(B,J) \arrow[r] & 0 . \end{tikzcd}$$

By examining the definition of ψ and ψ' we see that this commutes and an easy diagram chase yields Theorem 7.2.

We must now prove Corollary 7.3. If $x \in N$ and $y \in I$ then $xy = yx = 0$ so

$c(x, y) \in K_2(A)$ is defined. Its image in $K_2(A/I)$ is $c(0, y) = 1$ so $c(x, y) \in K'_2(A, I)$. Its image in $K'_2(B, J)$ is $c(x, 0) = 1$. To show that c induces the map ψ' of Corollary 7.3 we must make a few computations in $\mathrm{St}(A)$.

Lemma 7.4. *Let* $ab = ba = 0$. *Then for any pair of indices* $i \neq j$ *we have* $[x_{ij}(a), x_{ji}(b)] = c(a, b)$.

Proof. If $i = 1, j = 2$, this is the definition of $c(a, b)$. As in [9, §9] let $w_{kl} = x_{kl}(1)\, x_{lk}(-1)\, x_{kl}(1)$. As in [9, Cor. 9.4], an easy calculation with the Steinberg relations shows that $w_{kl}\, x_{ij}(r)\, w_{kl}^{-1} = w_{\pi(i)\pi(j)}(\epsilon r)$ where π is the transposition (kl) and $\epsilon = 1$ unless $i = k$ or $j = k$ when $\epsilon = -1$. If $j \neq 2$, we have $w_{j2}\, c(a, b)\, w_{j2}^{-1} = [x_{1j}(a), x_{j1}(b)]$. If $i \neq 1$ we have $w_{i1}\, [x_{1j}(a), x_{j1}(b)]\, w_{i1}^{-1} = [x_{ij}(a), x_{ji}(b)]$. Since $c(a, b)$ is central, the result follows from this.

Corollary 7.5. *If* $ab = ba = 0$, *then* $c(a, b) = c(b, a)^{-1}$.

This follows from the identity $[x, y] = [y, x]^{-1}$. The following result is a special case of [9, Lemma 8.1].

Lemma 7.6. *If* $a_\nu b = 0 = ba_\nu$ *for* $\nu = 1, 2$, *then* $c(a_1 + a_2, b) = c(a_1, b)\, c(a_2, b)$ *and* $c(b, a_1 + a_2) = c(b, a_1)\, c(b, a_2)$.

This follows from the identity $[xy, z] = x[y, z]x^{-1}[x, z]$ and the fact that $c(a, b)$ is central.

Lemma 7.7. *If* $xyz = zxy = yzx = 0$, *then* $c(x, yz)\, c(z, xy)\, c(y, zx) = 1$.

Proof. Since $c(a, -b) = c(a, b)^{-1}$ by Lemma 7.6, it will suffice to show that $c(x, -yz)\, c(z, -xy)\, c(y, -zx) = 1$. Using Lemma 7.4 and the Steinberg relations, we can write this expression as

$$[x_{12}(x), [x_{23}(y)^{-1}, x_{31}(z)]]\, [x_{31}(z), [x_{12}(x)^{-1}, x_{23}(y)]] \times$$

$$\times [x_{23}(y), [x_{31}(z)^{-1}, x_{12}(x)]] \; .$$

Since the c's are central, the fact that this is 1 follows immediately from P. Hall's identity

$${}^{\beta}[\alpha, [\beta^{-1}, \gamma]] \cdot {}^{\gamma}[\beta, [\gamma^{-1}, \alpha]] \cdot {}^{\alpha}[\gamma, [\alpha^{-1}, \beta]] = 1.$$

Here ${}^{\beta}x$ stands for $\beta x \beta^{-1}$, etc. Hall's identity is valid in any group and is proved by writing out all commutators and checking that everything cancels.

We can now show that the map ψ' of Corollary 7.3 is well-defined. By Lemma 7.6, there is a map $N\otimes I \to K'_2(A,I)$ given by sending $x\otimes y$ to $c(x,y)$. Since $xy=0$ for $x\in N$, $y\in I$, we have $c(ax,y)\,c(ya,x) = c(ax,y)\,c(ya,x)\,c(xy,a) = 1$ by Lemma 7.7. By Corollary 7.5, we get $c(ax,y) = c(x,ya)$. Similarly $c(xa,y) = c(x,ay)$. Therefore we get a map $N\otimes_{A^e} I \to K'_2(A,I)$. If $x_1, x_2 \in N$ then $c(x_1x_2,y) = c(x_1x_2,y) \times c(yx_1,x_2)\,c(x_2y,x_1) = 1$ by Lemma 7.7 since $IN = NI = 0$. Similarly $c(x,y_1y_2) = 1$ for $y_1,y_2\in I$. Thus $c(x,y)$ depends only on $x \bmod N^2$ and $y \bmod I^2$. Therefore the map ψ' of Corollary 7.3 is well-defined and it is clear that the composition of ψ' with $K'_2(A,I) \to K'_2(B,J)$ is zero.

Let $R \to R/I$ be a split epimorphism of rings with unit. Then there is a subring $\bar{R}$ of R with $R = \bar{R} \oplus I$. The extension $1 \to \mathrm{St}'(R,I) \to \mathrm{St}(R) \to \mathrm{St}(R/I) \to 1$ splits and we can identify $\mathrm{St}(\bar{R})$ with its image in $\mathrm{St}(R)$. It acts on $\mathrm{St}'(R,I)$ by conjugation so we can regard $\mathrm{St}'(R,I)$ as a group with $\mathrm{St}(\bar{R})$ acting as a group of automorphisms. I will refer to such a group as an $\mathrm{St}(\bar{R})$-group.

Lemma 7.8. *Let R be a ring with unit, $\bar{R}$ a subring with the same unit and I a 2-sided ideal of R with $R = \bar{R} \oplus I$. Then $\mathrm{St}'(R,I)$ is generated as an $\mathrm{St}(\bar{R})$ group by all $x_{ij}(y)$ with $y\in I$ and the following are a set of defining relations*

(1) $\quad x_{ij}(x+y) = x_{ij}(x)\,x_{ij}(y)\,.$

(2) $\quad [x_{ij}(x), x_{kl}(y)] = 1 \quad \text{if} \quad i\neq l,\ j\neq k\,,$

(3) $\quad [x_{ij}(x), x_{jk}(y)] = x_{ik}(xy) \quad \text{if} \quad i\neq k\,.$

(4) $\quad x_{ij}(r)\cdot x_{ij}(x) = x_{ij}(x) \quad \text{for} \quad r\in \bar{R}\,.$

(5) $\quad x_{ij}(r)\cdot x_{kl}(x) = x_{kl}(x) \quad \text{if} \quad i\neq l,\ j\neq k,\ r\in\bar{R}\,,$

(6) $\quad x_{ij}(r)\cdot x_{jk}(x) = x_{jk}(rx)\,x_{jk}(x) \quad \text{if} \quad i\neq k,\ r\in\bar{R}\,.$

(7) $\quad x_{ij}(r)\cdot x_{ki}(x) = x_{kj}(-xr)\,x_{ki}(x) \quad \text{if} \quad j\neq k,\ r\in\bar{R}\,.$

In relations (4) to (7), the dot indicates the operation of $\mathrm{St}(\bar{R})$ on $\mathrm{St}'(R,I)$. The proof is the same as that of Lemma 8.4 of [16].

Remark. For any 2-sided ideal I of a ring R, $\mathrm{St}(R,I)$ is defined to be $\mathrm{St}'(R_1,I_1)$ where R_1 is the pullback of the diagram

$$\begin{tikzcd} R_1 \arrow[r, "p_1"] \arrow[d, "p_2"'] & R \arrow[d] \\ R \arrow[r] & R/I \end{tikzcd}$$

and $I_1 = \ker p_1$. Since p_1 is split by the diagonal map $R \to R_1$ we can use Lemma 7.8 to get a presentation of $\mathrm{St}(R, I)$. It is generated as an $\mathrm{St}(R)$-group by the elements $y_{ij}(x) = x_{ij}((0, x))$ for $x \in I$. The relations are like those of Lemma 7.8 except that we write $y_{ij}(x)$ for $x_{ij}(x)$ and R for $\bar{R}$. This presentation was given in [14, p. 214] but it unfortunately appeared in a rather garbled form.

Now suppose we are in the situation of Theorem 7.2. If $s \in \mathrm{St}'(A, N)$ and $y \in I$, the element $[s, x_{ij}(y)]$ lies in $\mathrm{St}'(A, N) \cap \mathrm{St}'(A, I)$ since both these subgroups are normal in $\mathrm{St}(A)$. Therefore its images in $\mathrm{St}(A/N) = \mathrm{St}(B)$ and in $\mathrm{St}(A/I)$ are both trivial. The commutativity of

$$\begin{array}{ccc}
\mathrm{St}'(A, I) & \longrightarrow & \mathrm{St}'(B, J) \subset \mathrm{St}(B) \\
\varphi \downarrow & & \downarrow \varphi \\
\mathrm{GL}(A, I) & \xrightarrow{\approx} & \mathrm{GL}(B, J)
\end{array}$$

shows that $\varphi([s, x_{ij}(y)]) = 1$ so $[s, x_{ij}(y)]$ lies in $K_2'(A, I)$ and is central in $\mathrm{St}(A)$. Let C be the subgroup of $\mathrm{St}'(A, I)$ generated by all the elements $[s, x_{ij}(y)]$ for $s \in \mathrm{St}'(A, N)$, $y \in I$.

Lemma 7.9. *The sequence*

$$1 \to C \to \mathrm{St}'(A, I) \to \mathrm{St}'(B, J) \to 1$$

is exact.

Proof. If $t \in \mathrm{St}(A)$, $s \in \mathrm{St}'(A, N)$, $y \in I$, then $s(t\,x_{ij}(y)\,t^{-1})s^{-1} = t(t^{-1}st\,x_{ij}(y)(t^{-1}st)^{-1})t^{-1} \equiv t\,x_{ij}(y)\,t^{-1} \bmod C$ since $t^{-1}st \in \mathrm{St}'(A, N)$ and C is normal in $\mathrm{St}(A)$. Since the elements $t\,x_{ij}(y)\,t^{-1}$ generate $\mathrm{St}'(A, I)$, it follows that $\mathrm{St}'(A, N)$ acts trivially on $G = \mathrm{St}'(A, I)/C$.

Now let $\bar{A} = \alpha(A/I) \subset A$ and $\bar{B} = \beta(B/J) \subset B$. The commutativity of

$$\begin{array}{ccc}
\mathrm{St}(\bar{A}) & \longrightarrow & \mathrm{St}(A) \\
\downarrow & & \downarrow \\
\mathrm{St}(\bar{B}) & \longrightarrow & \mathrm{St}(B)
\end{array}$$

shows that

$$1 \to \mathrm{St}(\bar{A}) \cap \mathrm{St}'(A, I) \to \mathrm{St}(\bar{A}) \to \mathrm{St}(\bar{B}) \to 1$$

is exact, so we can regard G as an $\mathrm{St}(\bar{B})$-group. Define the map $\mathrm{St}'(B, J) \to G$ to be the $\mathrm{St}(\bar{B})$-homomorphism sending $x_{ij}(y)$ for $y \in J$ to the image in G of $x_{ij}(x)$ where $x \in I$, $f(x) = y$. The relations of Lemma 7.8 are easily verified. Since $\mathrm{St}'(B, J)$ is generated as an $\mathrm{St}(\bar{B})$-group by the $x_{ij}(y)$, and G is generated as an $\mathrm{St}(\bar{A})$-group (and

hence as an St($\bar{B}$)-group) by the images of the $x_{ij}(x)$, we see that the compositions of the two maps $G \to \mathrm{St}'(B,J)$ and $\mathrm{St}'(B,J) \to G$ in either order are the identity maps. Therefore $G \to \mathrm{St}'(B,J)$ is an isomorphism.

Corollary 7.10. *The sequence*

$$0 \to C \to K_2'(A,I) \to K_2'(B,J) \to 0$$

is exact.

This follows by the snake lemma applied to the diagram

$$\begin{array}{ccccccccc} 1 & \to & C & \to & \mathrm{St}'(A,I) & \to & \mathrm{St}'(B,J) & \to & 1 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 1 & \to & 1 & \to & \mathrm{GL}(A,I) & \to & \mathrm{GL}(B,J) & \to & 1\,. \end{array}$$

For any ring R with unit, let $\mathrm{St}(R)$ act on itself by inner automorphism. Since the center acts trivially, this action factors through the canonical map $\varphi\colon \mathrm{St}(R) \to E(R)$. If $\sigma \in E(R)$ and $x \in \mathrm{St}(R)$, we define $\sigma \cdot x = yxy^{-1}$ where y is any element of $\mathrm{St}(R)$ with $\varphi(y) = \sigma$.

Lemma 7.11. *Let $\sigma = (a_{\mu\nu}) \in E_n(R)$ and let $\sigma^{-1} = (b_{\mu\nu})$. Then*

(1) *If* $i,j > n$, $\sigma \cdot x_{ij}(r) = x_{ij}(r)$

(2) *If* $i > n$, $j \leq n$, $\sigma \cdot x_{ij}(r) = \prod_{\nu=1}^{n} x_{i\nu}(rb_{j\nu})$

(3) *If* $i \leq n$, $j > n$, $\sigma \cdot x_{ij}(r) = \prod_{\mu=1}^{n} x_{\mu j}(a_{\mu i} r)$.

Proof. This is clear for $\sigma = 1$. Since $E_n(R)$ is generated by all $e_{pq}(s)$ with $p, q \leq n$, $s \in R$, it will suffice to show that if the lemma is true for σ it is also true for $e_{pq}(s)\sigma$. Now $e_{pq}(s)\,\sigma \cdot x_{ij}(r) = x_{pq}(s)(\sigma \cdot x_{ij}(r))x_{pq}(s)^{-1}$. In case (1), $x_{pq}(s)$ and $x_{ij}(r)$ commute and the result is clear. In case (2), $x_{pq}(s)$ commutes with $x_{i\nu}(rb_{j\nu})$ for $\nu \neq p$ and $x_{pq}(s)\,x_{ip}(rb_{j\nu})\,x_{pq}(s)^{-1} = x_{iq}(-rb_{j\nu}s)\,x_{ip}(rb_{j\nu})$. Therefore, the only change in the right-hand side of (2) is that $x_{iq}(rb_{jq})$ is replaced by $x_{iq}(r(b_{jq} - b_{jp}s))$. But, if $(e_{pq}(s)\,\sigma)^{-1} = \sigma^{-1}e_{pq}(-s) = (b'_{\mu\nu})$ then $b'_{j\nu} = b_{j\nu}$ for $\nu \neq q$ and $b'_{jq} = b_{jq} - b_{jp}s$. The proof of (3) is similar.

We can now prove Corollary 7.3. The group C above is generated by all $[s, x_{ij}(y)]$ with $s \in \mathrm{St}'(A,N)$, $y \in I$. Let $\sigma = \varphi(s) \in E(A,N)$. Then $[s, x_{ij}(y)] = (\sigma \cdot x_{ij}(y))x_{ij}(y)^{-1}$. Suppose $\sigma \in E_n(A,N)$. Choose $N > n$. Then $x_{ij}(y) = [x_{iN}(y), x_{Nj}(1)]$ so $\sigma \cdot x_{ij}(y) = [\sigma \cdot x_{iN}(1), \sigma \cdot x_{Nj}(y)]$. Let $\sigma = (a_{\mu\nu})$ and $\sigma^{-1} = (b_{\mu\nu})$. Since $\sigma \in E_n(A,N) \subset \mathrm{GL}_n(A,N)$,

we have $a_{\mu\nu} - \delta_{\mu\nu} \in N$. Since $NI = 0$ and $y \in I$, $a_{\mu\nu}y = \delta_{\mu\nu}y$. By Lemma 7.11 we have $\sigma \cdot x_{iN}(y) = \Pi_{\mu=1}^{n} x_{\mu N}(a_{\mu i}y) = x_{iN}(y)$ and $\sigma \cdot x_{Nj}(1) = \Pi_{\nu=1}^{n} x_{N\nu}(b_{j\nu})$. Since $\sigma^{-1} \in \mathrm{GL}_n(A, N)$, $b_{j\nu} - \delta_{j\nu} \in N$ Therefore if $\nu \neq i$, $[x_{iN}(y), x_{N\nu}(b_{j\nu})] = x_{i\nu}(yb_{j\nu}) = 1$ unless $\nu = j$. It follows that $\sigma \cdot x_{ij}(y) = [x_{iN}(y), \Pi x_{N\nu}(b_{j\nu})] = [x_{iN}(y), x_{Nj}(b_{jj})x_{Ni}(b_{ji})] = [x_{iN}(y), x_{Nj}(b_{jj})] x_{Nj}(b_{jj}) \cdot [x_{iN}(y), x_{Ni}(b_{ji})] x_{Nj}(b_{jj})^{-1}$. The first bracket is $x_{ij}(yb_{jj}) = x_{ij}(y)$ since $b_{jj} - 1 \in N$. The second bracket is $c(y, b_{ji}) = c(b_{ji}, y)^{-1}$ by Lemma 7.4 and Corollary 7.5. Since this is central, $\sigma \cdot x_{ij}(y) = c(b_{ji}, y)^{-1} x_{ij}(y)$ so $[s, x_{ij}(y)] = c(b_{ji}, y)^{-1}$. Therefore the group C is generated by elements $c(b, y)$ with $b \in N$, $y \in I$ and so lies in the image of ψ'. This proves Corollary 7.3 and hence Theorems 7.1 and 7.2.

I will conclude by proving a stronger version of Lemma 7.11.

Theorem 7.12. *For any ring R with unit there is a unique action of* $\mathrm{GL}(R)$ *on* $\mathrm{St}(R)$ *which satisfies the identities of Lemma 7.11 for $\sigma \in \mathrm{GL}_n(R)$. This action extends the action of $E(R)$ on* $\mathrm{St}(R)$ *used in Lemma 7.11.*

Proof. If we have such an action and $\sigma \in \mathrm{GL}_n(R)$, choose $N > n$ and use the fact that $\sigma \cdot x_{ij}(r) = [\sigma \cdot x_{iN}(1), \sigma \cdot x_{Nj}(r)]$. The terms inside the brackets are determined by the identities of Lemma 7.11. Therefore the action is unique. If $\sigma \in E_n(R)$ the same argument together with Lemma 7.11 shows that the action agrees with the usual one.

Now let S_m be the subgroup of $\mathrm{St}(R)$ generated by all $x_{ij}(r)$ with $i, j \leqslant m$. If $\sigma \in \mathrm{GL}_n(R)$ and $x \in S_m$, let

$$\rho = \begin{pmatrix} \sigma & 0 & 0 \\ 0 & 1_{m-n} & 0 \\ 0 & 0 & \tau \end{pmatrix}$$

where τ is chosen to make $\rho \in E(R)$. For example, we may take $\tau = \sigma^{-1}$ [14, Lem. 13.3]. Define $\sigma \cdot x = \rho \cdot x$. If we choose a different τ_1 getting a new ρ_1, we have

$$\rho^{-1}\rho_1 = \begin{pmatrix} 1_m & 0 \\ 0 & \tau^{-1}\tau_1 \end{pmatrix}$$

and this acts trivially on x by Lemma 7.11(1) (we must renumber the indices here, of course). Therefore $\sigma \cdot x$ is defined. Clearly $\sigma \cdot (xy) = (\sigma \cdot x)(\sigma \cdot y)$. If $\sigma, \sigma' \in \mathrm{GL}_n(R)$ and we have found corresponding elements $\rho, \rho' \in E(R)$, we can choose $\rho\rho'$ to correspond to $\sigma\sigma'$. Therefore $\sigma\sigma' \cdot x = \sigma \cdot (\sigma' \cdot x)$.

Since we can choose τ to have the form

$$\tau = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^{-1} \end{pmatrix}$$

we see that the actions of $GL_n(R)$ on S_m and S_{m+1} agree. Therefore $GL_n(R)$ acts on $\cup S_m = St(R)$. The inclusions $GL_n(R) \subset GL_{n+1}(R)$ are clearly compatible with this action so $GL(R)$ acts on $St(R)$. The required identities follow directly from the definition of the action and Lemma 7.11.

References

[1] H. Bass, Algebraic K-theory (Benjamin, New York, 1968).

[2] H. Bass, J. Milnor and J.-P. Serre, Solution of the congruence subgroup problem for SL_n $(n \geq 3)$ and Sp_{2n} $(n \geq 2)$, Publ. Math. I.H.E.S. 33 (1967) 59–137.

[3] H. Bass and M.P. Murthy, Grothendieck groups and Picard groups of Abelian group rings, Ann. of Math. 86 (1967) 16–73.

[4] J. Dieudonné, Les déterminants sur un corps non-commutative, Bull. Soc. Math. France 71 (1943) 27–45.

[5] S.M. Gersten, On the functor K_2, I, J. Algebra 17 (1971) 212–237.

[6] S.M. Gersten, On Mayer-Vietoris functors and algebraic K-theory, J. Algebra 18 (1971) 51–88.

[7] N. Jacobson, Structure of rings, A.M.S. Colloq. Publ. 37, Providence 1956.

[8] M. Karoubi and O. Villamayor, Foncteurs K^n en algèbre et en topologie, Compt. Rend. Acad. Sci. Paris 269 (1969) 416–419.

[9] J. Milnor, Notes on algebraic K-theory, M.I.T. 1969 (mimeographed notes).

[10] D. Quillen, On the K-theory associated to a finite field, to appear.

[11] J.R. Silvester, A presentation of the GL_n of a semi-local ring, to appear.

[12] M.R. Stein, Relativizing functors on rings and algebraic K-theory, to appear.

[13] R. Steinberg, Lectures on Chevalley groups, Yale 1967 (mimeographed notes).

[14] R.G. Swan, Algebraic K-theory, Lecture Notes in Math. 76 (Springer-Verlag, Berlin, 1968).

[15] R.G. Swan, K-Theory of finite groups and orders, Notes by E.G. Evans, Lecture Notes in Math. 149 (Springer-Verlag, Berlin, 1970).

[16] R.G. Swan, Non-Abelian homological algebra and K-theory, Proc. Symp. Pure Math. A.M.S. 17 (1970) 88–123.

[17] R.G. Swan, Some relations between higher K-functors, J. Algebra 19 (1971), to appear.

[18] L.N. Wasserstein, On the stabilization of the general linear group over a ring, Mat. Sbornik 79 (1969) 405–424 (in Russian).

[19] O. Zariski and P. Samuel, Commutative algebra, Vol. II (Van Nostrand, Princeton, 1950).

JOURNAL OF PURE AND APPLIED ALGEBRA - Volume 1, No. 3 (1971) pp. 253-273.

SUBGROUPS OF FINITE DIMENSIONAL TOPOLOGICAL GROUPS

David L. RECTOR [1]
Rice University, Houston, Texas, USA

Received 3 February 1971

H-spaces which have the homotopy type of a finite CW-complex and, in particular, topological groups with that property share many of the homotopy theoretic properties of Lie groups. In particular, such spaces satisfy Poincaré duality [14, 7.9] and have rational cohomology rings which are exterior algebras on generators of odd degree [19]. Recently, the study of such H-spaces has been enriched by the construction, using a technique of A. Zabrodsky, of many examples which are not homotopy equivalent to any Lie group [34, 30, 18, 15].

Two sets of phenomena which have proven important in the theory of Lie groups are the maximal torus with its associated Weyl group and the properties of the cohomology ring of a homogeneous space. It is therefore interesting to know to what extent these phenomena generalize to the homotopy category of topological groups with the homotopy type of a finite complex (to be called FD-groups below). The generalization of these phenomena depends on giving a definition of subgroup which makes sense in the homotopy category. In §1 we propose such a definition which essentially requires that the quotient space of a group by a subgroup have the homotopy type of a finite complex. We will prove in §2 that quotient spaces are Poincaré complexes and that their cohomology rings satisfy various theorems of Borel et al., when the subgroup is of maximal rank. In §3 we examine the case of a subtorus and define an analogue of the Weyl group. We will show that this group is isomorphic to the classical Weyl group in case the FD-group and subtorus are a Lie group and a maximal subtorus in the classical sense.

In the remainder of the paper, we will provide examples of groups and subgroups, and, in particular, we will provide an example of a group which contains no torus of maximal rank. For the latter example we will use an exotic multiplication on S^3 first constructed by Slifker [28]. We will give a much simplified construction of these multiplications using Zabrodsky's homotopy mixing construction [34].

The author would like to thank Al Vasquez for a very useful discussion leading to the proof of Theorem 3.6. I would also like to thank Larry Smith and Frank Quinn for suggesting the central idea in the proof of Theorem 2.1, which is analogous to a theorem in the latters thesis (Princeton 1970).

[1] The author was partially supported by NSF-grant GP-20552 during the preparation of this work

§ 1. Subgroups

Results of Stasheff [29] et al., show that for homotopy theoretic purposes we may regard a topological group to be the loop space on some space and a homomorphism to be induced by a homotopy class of continuous mappings of spaces. For if G is a topological group, B_G its classifying space [25], then there is an ΩB_G fitted with a natural group structure and a homomorphism $\Omega B_G \to G$ which is an A_∞ equivalence in the sense of Stasheff. We will thus call two groups equivalent if they have homotopy equivalent classifying spaces.

The group structure on a loop space is technically easiest to construct in the appropriate simplicial category. If X is a simplicial set with one vertex, there is a free simplicial group GX and a principal fibration

$$GX \to PX \to X$$

with PX contractible [30, 22]. There is also for a simplicial group G a classifying fibration

$$G \to WG \to \overline{W}G$$

[26]. In the category of spaces, ΩX may be given a group structure by using a Moore loop space with appropriate reductions to insure the existence of inverses. Care must be taken to insure that all spaces, in particular product spaces, be given a compactly generated topology and are of the homotopy type of a CW-complex. The author is prepared to take the responsibility for details in the simplicial category and therefore the notation GX will be used for the loop complex of X. The more familiar notation B_G will be used for the classifying complex. Whenever an obvious space such as a Lie group is mentioned, the singular complex of that space will be meant. It will not be necessary for us to explicitly consider technical details in the simplicial categories.

We shall mean by an *FD-group structure* a simplicial set X with a single vertex such that GX has the homotopy type of a finite complex. A *homomorphism* is a simplicial map $X \to Y$ of FD-group structures. Two homomorphisms are equivalent if they are homotopic and two group structures are equivalent if they are homotopy equivalent. If K is a finite complex, we shall mean by a *group structure on K* an FD-group structure X such that $GX \simeq K$. If X is an equivalence class of FD-group structures we will call $H = GX$ an *FD-group*. Now, given a simplicial group H, we have a group structure B_H functorially determined. This structure is unique up to equivalence since there is a natural homotopy equivalence $X \simeq B_{GX}$. So if H is an FD-group, we denote by B_H the FD-group structure determining it. A similar convention will be applied to homomorphisms.

Let $H \xrightarrow{f} G$ be a homomorphism of FD-groups. By replacing B_G by a homotopy equivalent complex $\widetilde{B}_G$ (mapping cylinder) we may replace the map B_f by an equivalent inclusion. Thus we may replace the homomorphism f by an equivalent inclusion of simplicial groups $GB_H \to G\widetilde{B}_G$. We then have a principal fibre bundle [8]

$$GB_H \to G\widetilde{B}_G \to G\widetilde{B}_G/GB_H.$$

The homotopy type of the base of this fibre bundle is easily seen to be determined by the equivalence class of f. We denote the base by $G//f$. We now propose

1.1. Definition. A homomorphism $f: H \to G$ of FD-groups is an *inclusion of FD-groups* if $G//f$ has the homotopy type of a finite complex. In that case we denote $G//f$ by G/H and the principal fibration above by

$$H \xrightarrow{f} G \to G/H \ .$$

An alternate definition could be obtained as follows. Replace B_H by an equivalent complex $\tilde{B}_H$ and B_f by an equivalent map $\tilde{B}_f$ so that $\tilde{B}_f$ is a Kan fibration with fibre F. Then $F \simeq G//f$. So

1.2. Proposition. *$f: H \to G$ is an inclusion of FD-groups iff the fibre of B_f (made into a fibration) has the homotopy type of a finite complex.*

Our notion of inclusion is a generalization of the classical notion for Lie groups in the following sense. Let $H \to G$ be a pair of compact Lie groups. Then G/H is a compact manifold and is homotopy equivalent to the fibre of the inclusion of classifying spaces $B_H \to B_G$; thus

1.3. Proposition. *If $H \to G$ is an inclusion of Lie groups it is an inclusion of FD-groups.*

We have the following transitivity property of FD-inclusion.

1.4. Proposition. *Let $K \subseteq H$ and $H \subseteq G$ be FD-inclusions. Then the composition $K \subseteq G$ is an FD-inclusion.*

Proof. We have fibrations

$$F_0 \to B_H \to B_G$$

$$F_1 \to B_K \to B_H$$

with F_0 and F_1 having the homotopy type of finite complexes. Let F_{01} be the fibre of the composition $B_K \to B_H \to B_G$. Then there is a fibration

$$F_1 \to F_{01} \to F_0 \ .$$

By [23; Th. 1] the Wall obstruction to finiteness of F_{01} [33] vanishes.

Now suppose $F \to X \xrightarrow{f} Y$ is a fibration. Then the sequence of simplicial groups

$$GF \to GX \xrightarrow{Gf} GY$$

is not an exact sequence. But if $\tilde{G}F = \operatorname{Ker} Gf$, then there is an inclusion of free simplicial groups $GF \subset \operatorname{Ker} Gf$ which induces isomorphisms of homotopy groups. Thus $B_{GF} \simeq B_{\tilde{G}F} \simeq F$. So we have an exact sequence of groups

$$\tilde{G}F \to GX \to GY$$

with $\tilde{G}F$ equivalent to GF. Thus we may make

1.5. Definition. A sequence of FD-groups $K \to H \to G$ is an *extension* if it is equivalent to an exact sequence of groups, i.e., if the sequence $B_K \to B_H \to B_G$ is a quasifibration [17]. An FD-inclusion $K \to H$ is *normal* if it is equivalent to the fibre of an extension.

Let K and G be FD-groups. We wish to classify all extensions $K \to H \to G$. This is equivalent to classifying fibre homotopy equivalent fibrations

$$B_K \to B_H \to B_G.$$

Let $H(B_K)$ be the complex of homotopy equivalences of B_K (see Appendix). Then fibre homotopy classes of fibrations $B_K \to X \to B_G$ are in one to one correspondence with the elements of $[B_G, H(B_K)]$ (see Appendix). Thus

1.6. Proposition. *Let K and G be FD-groups. The equivalence classes of extensions of FD-groups $K \to H \to G$ correspond one for one to elements of $[B_G, H(B_K)]$.*

Proof. We have yet to show that if

$$B_K \to X \to B_G$$

is a fibration then X is an FD-group structure. But there is a fibration $F \to GX \to GB_G$, $F \simeq GB_K$ with GB_K and GB_G having the homotopy type of finite complexes. Thus, again by [23, Th. 1], GX has the homotopy type of a finite complex.

§ 2. FD-homogeneous spaces

We will show in this section that quotients G/H of a pair of FD-groups $H \subseteq G$ share many of the homotopy theoretic properties of homogeneous spaces of Lie groups.

2.1. Theorem. *Let G and H be connected FD-groups, H an FD-subgroup of G. Then G/H is a Poincaré complex.*

Proof. Let $\dim H = m$, $\dim G = n$, $\dim G/H = l$. Consider the principal fibration

$$H \to G \to G/H.$$

Since H and G are connected groups, $\pi_1 G/H$ acts trivially on all homology groups in sight. Let $\{E^r\}$ and $\{E_r\}$ be the homology and cohomology Serre spectral sequences of the fibration above with coefficients in the commutative ring Λ. Since $\dim H = m$, $\dim G/H = l$, it is clear from the spectral sequence $\{E^r\}$ that $\dim G = l + m$. In fact, since $H_m(H; Z) = Z$, $H_n(G; Z) = Z$, and no nontrivial differentials enter

or leave $E^r_{l,m}$, we have

$$H_l(G/H;Z) = E^2_{l,m} = H_n(G;Z) = Z\,.$$

The basic idea of our proof is this. We want to construct a map of spectral sequences

$$\{f_r : E_r^{p,q} \to E^r_{l-p,m-q}\}$$

such that

(i) $f_2 : E_2^{0,q} \to E^2_{l,m-q} \approx H_{m-q}(H;\Lambda)$ is the duality map $H^q(H;\Lambda) \to H_{m-q}(H;\Lambda)$.

(ii) $f_2 : E_2^{p,0} \to E^2_{l-p,m} \approx H_{l-p}(G/H;\Lambda)$ is cap product with the top class.

(iii) The duality map $H^i(G;\Lambda) \to H_{n-i}(G;\Lambda)$ preserves filtrations from the spectral sequence, and $f_\infty : E_\infty \to E^\infty$ is the graded map associated to the duality map.

If these conditions are satisfied, then with appropriate reindexing of $\{F^r\}$, the maps f_r satisfy the conditions of the Zeeman comparison theorem [35]. Thus the cap product $H^p(G/H;\Lambda) \to H_{l-p}(G/H;\Lambda)$ will be an isomorphism. To carry out this program without going to the trouble of investigating cap products in $\{E_r\}$ we proceed as follows.

First we suppose Λ to be a field k. For each finite complex X we may identify $H_*(X;k)$ with $(H^*(X;k))^*$. Also we may identify $E^r_{p,q}$ with $(E_r^{p,q})^*$. The filtration of $H_*(G;k)$ from the spectral sequence may now be identified as

$$F^r H_*(G;k) = \{f \in (H^*(G;k))^* \mid f(F_{r+1}H^*(G;k)) = 0\}\ .$$

Duality arises as follows. We have a graded symmetric bilinear form $\langle\ ,\ \rangle$: $H^*(G,k) \otimes H^*(G,k) \to k$ given by

$$\langle x,y\rangle = \begin{cases} xy, & \dim x + \dim y = n\,, \\ 0 & \text{otherwise.} \end{cases}$$

The duality map of G is $f: H^*(G;k) \to H_{n-*}(G;k)$ given by

$$f(x)(y) = \langle x,y\rangle\ .$$

This map is an isomorphism. Now in $H^*(G;k)$,

$$F_p \cdot F_q = 0\,, \quad p+q > l$$

since $E^2_{s,*} = 0$, $s > l$. Therefore, $f(F_p) \subseteq F^{l-p}$. Similarly, define a form $\langle\ ,\ \rangle_r$ on E_r by

$$\langle x,y\rangle_r = \begin{cases} x\cdot y\,, & \operatorname{bideg} x + \operatorname{bideg} y = (l,m), \\ 0\,, & \text{otherwise.} \end{cases}$$

We have f_r given by

$$f_r(x)(y) = \langle x, y \rangle_r \ .$$

The maps f_r satisfy the conditions (i), (ii), (iii) above. All that remains is to show $f_r \circ d_r = d^r \circ f_r$. But if bideg x + bideg $d_r y = (l, m)$, then $f_r(x)(d_r y) = x \cdot d_r y = \pm d_r x \cdot y = f_r(d_r x)(y)$ since d_r is a derivation. We have thus proven the theorem.

We will now give detailed results for $H^*(G/H; k)$ when H is of maximal rank. Recall [19] that if G is any connected FD-H space, then

$$H^*(G; Q) = \Lambda_Q [x_{2r_1 - 1}, \ldots, x_{2r_n - 1}] \ ,$$

an exterior algebra on n generators of odd degree. The number n is an invariant of G and we call it rank G. If H is an FD-subgroup of G, we say H is of *maximal rank* if rank H = rank G.

Let k be a field, p = char k. Suppose $p = 0$, or $H^*(G; Z)$ has no p-torsion; then

$$H^*(G; k) = \Lambda_k [x_{2r_1 - 1}, \ldots, x_{2r_n - 1}] \ ,$$

and, by a standard Serre spectral sequence argument,

$$H^*(B_G; k) = k[y_{2r_1}, \ldots, y_{2r_n}] \ ,$$

where the generators $x_{2r_i - 1}$ transgress to y_{2r_i}.

Many of the classical theorems on cohomology of homogeneous spaces (see e.g. [12]) are included in

2.2. Theorem (Borel [11]). *Let H be a connected FD-subgroup of the connected FD-group G, k a field of characteristic p. Suppose*

(a) *H and G have no p-torsion or $p = 0$,*

(b) *H is of maximal rank.*

Then, if $f: B_H \to B_G$ is the inclusion,

(i) *$H^*(B_G; k) \xrightarrow{f^*} H^*(B_H; k)$ is monic,*

(ii) *$H^*(G/H; k) = H^*(B_H; k)//H^*(B_G; k)$ as an algebra* (i.e. $H^*(G/H; k) = H^*(B_H; k) \otimes_{H^*(B_G; k)} k$),

(iii) *$H^*(B_H; k) \approx H^*(G/H; k) \otimes_k H^*(B_G; k)$ as a $H^*(B_G; k)$ module, i.e. $H^*(B_H; k)$ is a free $H^*(B_G; k)$ module of dimension equal to $\chi(G/H)$, where χ is the Euler characteristic.*

It is an open question whether the attempted generalization in the Lie case by Baum [9, 10] to subgroups of less than maximal rank holds for FD-groups. The result for Lie groups has been announced by P. May.

Proof of 2.2. The proof by Baum [10] goes through without change. We outline the proof here since we wish to draw some corollaries from the proof. The essential ingredients are the following two lemmas. Let $\Lambda = k[x_1, \ldots, x_n]$, $\deg x_i > 0$, and let $I \subset \Lambda$ be an ideal, $I = (y_1, \ldots, y_m)$, where the y_i are homogeneous elements of Λ and are an irredundant set of generators of I. Then [10]

2.3. Lemma. *$\dim_k \Lambda/I$ is finite iff $m \geq n$.*

2.4. Lemma. *If $\dim_k \Lambda/I$ is finite, I is a Borel ideal iff $m = n$.*

The theorem now follows from these lemmas and a spectral sequence argument on the fibration

$$G/H \to B_H \to B_G.$$

From 2.3 we have immediately,

2.5. Corollary. *If H is an FD-subgroup of G, then rank $H \leq$ rank G.*

§3. Tori and the Weyl group

Let T be a connected homotopy Abelian FD-group. Then by a theorem of Hubbuck [20], T has the homotopy type of a product of circles. But then B_T is a product of $K(\mathbf{Z}, 2)$'s and is unique up to homotopy type. Therefore

3.1. Proposition. *There is a unique torus T^n with* rank $T^n = \dim T^n = n$.

Corollary 2.5 becomes

3.2. Proposition. *If G is an FD-group and T a subtorus, then* rank $T \leq$ rank G.

An FD-group need not have a torus of maximal rank. In §3 and §4 we will provide an example.

Let $B_T \xrightarrow{f} B_G$ be a subtorus. We have the following analogue of the Weyl group of a Lie group.

3.3. Definition. The Weyl group of G with respect to the inclusion of a subtorus $f\colon B_T \to B_G$ is

$$W(G, f) = \{\alpha \in [B_T, B_T] \mid \alpha \text{ is a homotopy equivalence and } f \circ \alpha \sim f\}.$$

Note that since B_T is an Eilenberg-MacLane space, $[B_T, B_T]$ is isomorphic to $\mathrm{Hom}(\mathbf{Z}^n, \mathbf{Z}^n)$, so

(3.4) $\qquad W(G, f) \subseteq \mathrm{Gl}(n, \mathbf{Z}),\quad \text{rank } T = n.$

3.5. Remark. We may generalize W as follows. Let $H(G, f)$ be the complex of all homotopy equivalences α: $B_T \to B_T$ (see Appendix) such that $f \circ \alpha \sim f$. Let

$$W_*(G, f) = \pi_* H(G, f).$$

Clearly $W_0(G, f) = W(G, f)$. Perhaps these groups will provide useful information about G if they can be calculated. It is plausible that $W_i(G, f) = 0$, $i > 0$, if G is a Lie group and f the classifying map of a Lie maximal torus.

We are justified in calling W a Weyl group since

3.6. Theorem. *Let G be a connected Lie group, T^n a maximal torus (in the Lie sense), and W the Weyl group of G. Then*

$$W \approx W(G, f)$$

where f is the classifying map of the inclusion $T \to G$.

Proof. Let $\alpha \in W$. Then α induces a nontrivial automorphism $H^1(T^n, Z) \to H^1(T^n, Z)$, so α may be considered to be an element of $\mathrm{Gl}(n, Z)$. To see that $\alpha \in W(G, f)$, recall that α is induced by an inner automorphism $a^{-1}(\ \ )a$ sending T into T. Let h: $a \sim 1$ be a path from a to 1. Then define H: $T \times I \to G$ by

$$H(x, t) = h(t)^{-1} x\, h(t)\,.$$

H is a continuous deformation by homomorphisms of $a^{-1}(\ \ )a$ to the inclusion $T^n \to G$. Thus $a^{-1}(\ \ )a$ and the inclusion induce homotopic maps $B_T \to B_G$. Thus $W \subseteq W(G, f)$. We must now show $W(G, f) \subseteq W$. Let $\Lambda = H^*(B_T; Q)$. Then $\Lambda = Q[x_1, ..., x_n]$, $\dim x_i = 2$. Now $H^*(B_G; Q)$ may be considered to be a subalgebra Γ of Λ, and $\Gamma = \Lambda^W$, the fixed subalgebra under the action of W. Clearly Γ is also fixed under the action of $W(G, f)$. To show $W(G, f) \subseteq W$ we will use some elementary Galois theory.

Let K be the quotient field of Λ, K^W the fixed subfield under the action of W, and $K_0 \subseteq K^W \subseteq K$ the quotient field of Γ. Now $K : K^W$ is a normal extension [5, Cor. to Th. 14]; so it suffices to show K^W is fixed under $W(G, f)$. But $W(G, f)$ leaves K_0 fixed so it suffices to show $K_0 = K^W$. Now $[K : K^W] = |W|$, so it suffices to show $[K : K_0] = |W|$. We recall from Lie theory that Λ is a free Γ module and $\dim_\Gamma \Lambda = |W|$ [11, 27.1]. Since Λ is a finite Γ module, Λ is integral over Γ [32, pp. 74–76]. Furthermore, since every rational function which is integral over the polynomial ring Λ is in Λ, Λ is the integral closure of Γ in K. But then [32, p. 78] any $x \in K$ may be written $x = a/b$ where $a \in \Lambda$ and $b \in \Gamma$. Thus $[K : K_0] = \dim_\Gamma \Lambda = |W|$ and $K_0 = K^W$.

3.7. Remark. We see from the above proof that the analogue for FD-groups of the classical theorem that any two maximal tori of a Lie group are conjugate would be that any two inclusions f, g: $B_T \to B_G$ of a torus T of maximal rank be homotopic up to a self equivalence of B_T. We do not as yet have such a theorem. Indeed, if G is a Lie group, we do not know whether an inclusion of a torus is homotopic to the

classifying map of a Lie inclusion. Thus the Weyl group defined above must *a priori* be taken to depend on the inclusion f.

§4. Tori in groups on S^3

We here undertake to provide an example of an FD-group containing no non-trivial torus. We will use an exotic structure on S^3. To distinguish between group structures on S^3 and to refute the existence of subtori in appropriate cases, we will use Adams operations in KO-theory.

Let B be a group structure on S^3, i.e., $GB \simeq S^3$. By using the Serre spectral sequence of the fibration

$$GB \to PB \to B$$

we may show that

$$H^*(B;Z) = Z[x_4],$$

$\dim x_4 = 4$. We want to calculate the KO-theory of B. For a systematic account of KO-theory see [13], or [3]. We need the following facts:

$$\mathrm{KO}^i(pt) = \begin{cases} Z, & i \equiv 0 \text{ or } -4 \pmod 8 \\ Z_2, & i \equiv -1 \text{ or } -2 \pmod 8 \\ 0, & \text{otherwise;} \end{cases}$$

$$\mathrm{KU}^i(pt) = \begin{cases} Z, & i \equiv 0 \pmod 2 \\ 0, & \text{otherwise.} \end{cases}$$

We have a generator η of $\mathrm{KU}^{-2}(pt)$ so that $\mathrm{KU}^*(pt) = Z[\eta, \eta^{-1}]$, and generators $\eta_1, \eta_2, \eta_4, \eta_8, \eta_i \in \mathrm{KO}^{-i}(pt)$ so that $\mathrm{KO}^*(pt)$ is generated by $\eta_1, \eta_4, \eta_8, \eta_8^{-1}$ as an algebra with relations $2\eta_1 = 0, \eta_1^2 = \eta_2, \eta_1\eta_2 = 0, \eta_1\eta_4 = 0, \eta_4^2 = 4\eta_8$. Under the complexification homomorphism c: $\mathrm{KO}^*(X) \to \mathrm{KU}^*(X)$, $c\eta_4 = 2\eta^2$, $c\eta_8 = \eta^4$.

We may now use the Atiyah-Hirzebruch spectral sequence for B [3, 7] to calculate $\mathrm{KO}^*(B)$ and $\mathrm{KU}^*(B)$.

4.1. Lemma. $\mathrm{KO}^*(B) = \mathrm{KO}^*(pt)[[x]]$, $x \in \mathrm{KO}^4(B)$, *and x represents $x_4 \in E_2^{4,0}(B)$ in the Atiyah-Hirzebruch spectral sequence.* $\mathrm{KU}^*(B) = \mathrm{KU}^*[[\bar{x}]]$, $\bar{x} \in \mathrm{KU}^4(B)$, $\bar{x}$ *representing x_4, and $\bar{x}$ may be chosen so that $cx = \bar{x}$.*

Proof. The spectral sequences collapses for dimensional reasons.

In order to distinguish homotopy types of the spaces B we are going to use the Adams operation Ψ^2 [3] applied to $\eta_4 x \in \mathrm{KO}(B)$. To do this we must know how this operation depends on the choice of the generator x. We will only be concerned

with this operation evaluated on $\eta_8 x^2$ so we will do our computations in $\mathrm{KO}^*/\mathrm{KO}_9^*$.

Note that an additive basis for $\mathrm{KO}^4(B)/\mathrm{KO}_9^4(B)$ is $\{x, \eta_4 x^2\}$. If x' is another representative of x_4, then $x' = x + e\eta_4 x^2$. Put $y = \eta_4 x$, $y' = \eta_4 x'$, $z = \eta_8 x^2$, $\bar{y} = \eta^2 \bar{x}$. Then $y' = \eta_4 x' = \eta_4 x + e\eta_4^2 x^2 = y + 4ez$. Now since $y \in \mathrm{KO}_4$, $\Psi^2 y = 4y + a'z$ [3, 5.2]. We assert a' is divisible by two. For $cy = 2\bar{y} \in \mathrm{KU}(B)$ and $cz = \bar{y}^2$, so $c\Psi^2 y = 2(\Psi^2 \bar{y}) = 8\bar{y} + 2a\bar{y}^2 = c(4y + 2az)$. Thus

$$\Psi^2 y = 4y + 2az \pmod{\mathrm{KO}_9}.$$

We must determine how a depends on the choice of x. Calculate modulo $\mathrm{KO}_9(B)$ to get

$$\begin{aligned}\Psi^2 y' &= \Psi^2(y + 4ez)\\ &= 4y + 64ez + 2az\\ &= 4y' + (2a + 48e)z\\ &= 4y' + 2(a + 24e)z\,.\end{aligned}$$

Thus a is well determined (mod 24). The choice of x_4 was unique up to sign. Also $\Psi^2(-y) = 4(-y) - 2az$. Denote a by $a(B)$. We have proved

4.2. Lemma. *If $B \simeq B'$, then $a(B) \equiv \pm a(B') \pmod{24}$*

We will prove in the next section that

4.3. Theorem. *There exist FD-group structures B on S^3 such that*

$$a(B) \equiv \pm 1, \pm 5, \pm 7, \quad \text{or} \quad \pm 11$$

(mod 24).

4.4. Remark. Note that $\mathrm{KO}(B)/\mathrm{KO}_9(B)$ is just the KO-theory of the projective plane of the multiplication on S^2 determined by B. The invariant $\pm a(B)$ distinguishes the four H-classes of multiplications on S^3 which may be made associative [28].

The main result of this section is

4.5. Theorem. *If B is a group structure on S^3 with $a(B) = \pm 5$ or ± 7, then B contains no non-trivial torus. There is a B with $a(B) = \pm 1$ which contains a circle.*

Proof. The B with $a(B) = \pm 1$ maybe taken to be HP^∞ by 4.6 below. The case $a(B) = \pm 11$ is as yet undetermined. Let B be a group structure on S^3, x_4 the generator of $H^4(B; \mathbf{Z})$ and $f\colon CP^\infty \to B$ a subtorus of rank one. Now $H^*(CP^\infty; \mathbf{Z}) = \mathbf{Z}[t]$, $\dim t = 2$. An immediate consequence of Theorem 2.2 is that $f^*(x_4) = \pm t^2$. When $a(B) = \pm 5$ or ± 7, we will show f does not exist by showing that $f^!$ cannot commute with Ψ^2. We get no more information from $\mathrm{KO}(CP^\infty)$ than from $\mathrm{KU}(CP^\infty)$ so we will calculate in KU-theory.

Recall that $\mathrm{KU}(CP^\infty) = Z[[\xi]]$, where $\xi + 1$ is the canonical line bundle, and ξ represents t in the Atiyah-Hirzebruch spectral sequence. Furthermore

$$\Psi^k \xi = (\xi + 1)^k - 1$$

[3]. Chose x_4 so that $f^*(x_4) = t^2$. We have $\mathrm{KU}(B) = Z[[\bar{y}]]$, y representing x_4. Therefore

$$f^! \bar{y} \equiv \xi^2 + e_3 \xi^3 + e_4 \xi^4 \pmod{\mathrm{KU}_9}.$$

Furthermore

$$\Psi^2 \bar{y} \equiv 4\bar{y} + a\bar{y}^2 \pmod{\mathrm{KU}_9}.$$

Calculate $\pmod{\mathrm{KU}_9}$:

$$\begin{aligned}\Psi^2 f^! \bar{y} &\equiv \Psi^2(\xi^2 + e_3\xi^3 + e_4\xi^4) \\ &\equiv 4\xi^2 + (4 + 8e_3)\xi^3 + (1 + 12e_3 + 16e_4)\xi^4 ,\end{aligned}$$

$$\begin{aligned}f^! \Psi^2 y &\equiv f^!(4\bar{y} + a\bar{y}^2) \\ &\equiv 4\xi^2 + 4e_3\xi^3 + (a + 4e_4)\xi^4 .\end{aligned}$$

Equating coefficients gives $e_3 = -1$ and

$$a = (1 + 12(e_4 - 1)).$$

Therefore

$$a \equiv 1 \pmod{12}.$$

The theorem is thus proved.

4.6. Lemma. $a(HP^\infty) = 1$.

Proof. If $B = HP^\infty$ in Lemma 4.1, the generator x may be chosen so that $\eta_8 x + 1_H = \lambda$, the canonical quaternionic line bundle [31, p. 250]. Let $f\colon CP^\infty \to HP^\infty$ be the inclusion and let $\zeta = \xi + 1$ be the canonical line bundle over CP^∞. Then

$$f^! \lambda = \zeta \oplus \bar{\zeta}$$

as a complex bundle [31, p. 251]. Therefore as a complex virtual bundle

$$\begin{aligned}\eta^{-2} c f^! \eta_8 x &= 2 - \zeta \oplus \bar{\zeta} \\ &= 1 - \xi - \frac{1}{1+\xi} \\ &= \xi^2/(1+\xi)\end{aligned}$$

since $\bar{\zeta} = 1/\zeta$ [3, 5.1]. Therefore

$$\Psi^2 \eta^{-2} c f^! \eta_8 x = (4\xi^2 + 4\xi^3 + \xi^4)/(1+\xi)^2$$
$$= 4\xi^2/(1+\xi) + \xi^4/(1+\xi)^2 .$$

Consequently,

$$\Psi^2 \eta_4 x = 4\eta_4 x + 2\eta_8 x^2$$

since $f^!$ and c are monic [4], [1, §2].

§ 5. Multiplications on S^3

We will now provide the multiplications of Theorem 4.3. Suitable examples were first constructed by Slifker [28] using other methods. We give here a much simplified construction using the Zabrodsky mixing procedure [34]. We will use this construction in the next section to find sub three-spheres in an exotic FD-group.

Let

$$f_i^! : HP^\infty \to K(Z, 4), \quad i = 0, 1,$$

be maps such that

(i) $f_i^!$ represents n_i times a chosen generator u_4 of $H^4(HP^\infty; Z)$, where

(ii) n_0 is odd, and

(iii) n_1 is a power of two.

By a theorem of Zabrodsky [34, 1.2], we may factor $f_i^!$ by

$$HP^\infty \xrightarrow{f_i} M_i \xrightarrow{f_i''} K(Z, 4),$$

where

(i) f_i'' is a fibre map, $i = 0, 1$,
(ii) f_0 and f_1'' are mod 2 equivalences [27],
(iii) f_0'' and f_1 are mod 3, 5, 7, ... equivalences,
(iv) all maps f_i, f_i'' are rational equivalences,
(v) $f_i^*: H^4(M_i; Z) \to H^4(HP^\infty; Z)$ has degree n_i, $i = 0, 1$.

Let B be the fibre product of f_0'' and f_1''. We have a commuting diagram

$$
\begin{tikzcd}
HP^\infty \arrow[r, "f_0"', "2\sim"] & M_0 \arrow[r, "f_0''", "3{,}5{,}7{,}\ldots\sim"'] & K(Z,4) \\
 & B \arrow[u, "\rho_0"] \arrow[r, "\rho_1"] & M_1 \arrow[u, "f_1''"', "2\sim"] \\
 & & HP^\infty \arrow[u, "f_1"', "3{,}5{,}7{,}\ldots\sim"]
\end{tikzcd}
\tag{5.1}
$$

It is clear from a Serre spectral sequence argument that ρ_0 is a 2-equivalence and ρ_1 is a 3,5,7, ...-equivalence. Thus

5.2. Lemma. $H^*(B; Z) = Z[x_4]$, $\dim x_4 = 4$, *and* $GB \simeq S^3$.

5.3. Proposition. $\pm a(B) \equiv 9n_0 - 8n_1 \pmod{24}$.

That this implies Theorem 4.3 is evident from the following table

n_0 (mod 8)	n_1 (mod 3)	$\pm a(B)$ (mod 24)
1	1	± 1
1	-1	± 7
3	1	± 5
3	-1	± 11

Proof of 5.3. Carry over the notation of §4 for KO(B). Let KO$^*(HP^\infty)$ = KO$^*(pt)[[u]]$, and let $v = \eta_4 u$, $w = \eta_8 u^2$. By Lemma 4.6 we may choose u so that

$$\Psi^2 v = 4v + 2w\,.$$

We will compute $a(B)$ by chasing diagram 5.1 with KO-theory. Since M_0 and M_1 have very bad torsion, we must localize KO-theory at the relevant primes. So let

$$
\begin{aligned}
{}_{1/2}\mathrm{KO}^*(X) &= \lim_{\substack{\leftarrow \\ n}} \mathrm{KO}^*(X^n) \otimes Z[\tfrac{1}{2}] \\
{}_2\mathrm{KO}^*(X) &= \lim_{\substack{\leftarrow \\ n}} \mathrm{KO}^*(X^n) \otimes Z_{(2)}\,,
\end{aligned}
$$

where $Z_{(2)}$ is the integers localized at the prime ideal (2). These theories are cohomology theories with ring and Ψ^k structures.

We may chose generators $x_0 \in {}_2\mathrm{KO}^4(M_0)$ and $x_1 \in {}_{1/2}\mathrm{KO}^4(M_1)$ so that $\rho_i^! x_i = x$, where we will use the symbol x also for the inclusion of x in ${}_{1/2}\mathrm{KO}$ or ${}_2\mathrm{KO}$. Then

$$ {}_2\mathrm{KO}(M_0) = {}_2\mathrm{KO}(pt)[[x_0]] $$

$$ {}_{1/2}\mathrm{KO}(M_1) = {}_{1/2}\mathrm{KO}(pt)[[x_1]] . $$

Let $y_i = \eta_4 x_i$, $z_i = \eta_8 x_i^2$, so

$$ \Psi^2 y_i \equiv 4y_i + 2az_i \pmod{\mathrm{KO}_9}. $$

Now

$$ f_i^! x_i = n_i u + e_i \eta_4 u^2 \pmod{\mathrm{KO}_9} $$

so

$$ f_i^! y_i = n_i v + 4e_i w . $$

Also

$$ \Psi^2 f_i^! y_i = 4n_i u + (2n_i + 64e_i) w $$

$$ f_i^! \Psi^2 y_i = 4n_i v + (2n_i^2 a + 16e_i) w . $$

So, recalling that the denominator of e_0 is prime to 2 and that e_i is a power of 2, we have

$$ a \equiv 1/n_0 \pmod 8 $$

$$ a \equiv 1/n_1 \pmod 3 . $$

But $1/n_0 \equiv n_0 \pmod 8$, $1/n_1 \equiv n_1 \pmod 3$ so

$$ a \equiv 9n_0 - 8n_1 \pmod{24}. $$

§6. The Hilton-Roitberg example

To illustrate how the techniques of §4 and §5 may be used to construct and study subgroups of exotic FD-groups, we will show that for each BS^3 constructed in §5, there is a multiplication on the Hilton-Roitberg "criminal" [18, 30] containing that BS^3. Associative multiplications on the Hilton-Roitberg bundle were first constructed by Stasheff [30].

We shall consider principal S^3 bundles over S^7. We will distinguish homotopy types of total spaces by a KO-theory invariant.

Let E be a simply connected complex such that $H^*(E; Z) = \Lambda_Z[t_3, t_7]$, $\dim t_i = i$. The total space of an S^3 fibration over S^7 is such a space. Using the Atiyah-Hirzebruch spectral sequence,

$$\mathrm{KO}^*(E) = \Lambda_{\mathrm{KO}^*(pt)}[t_3, t_7] ,$$

where t_3 and t_7 denote representatives of the corresponding elements in $H^*(E; Z)$. We want to compute Ψ^2 in $\mathrm{KO}^{-1}(E) = \mathrm{KO}(SE)$. Put $u = \eta_4 t_3$, $w = \eta_8 t_7$, then

$$\Psi^2 u = 4u + 4bw \quad (\mathrm{mod}\, \mathrm{KO}_9) ,$$

using, as in §4, complexification and the fact that $\Psi^2 u \equiv u^2 \pmod 2$ [6, 3.2.2]. If t_3' is another choice for t_3, $t_3' = t_3 + e\eta_4 t_7$, then for $u' = \eta_4 t_3'$, $u' = u + 4ew$. As in §4 we compute

6.1. Lemma. *$\pm b(E)$ is a homotopy invariant (mod 12).*

Let $S^3 \to E \to S^7$ be a fibration, and let $f\colon S^7 \to S^7$ be of degree d. Let E_f be the fibre product

$$\begin{array}{ccc} S^3 & = & S^3 \\ \downarrow & & \downarrow \\ E_f & \to & E \\ \downarrow & & \downarrow \\ S^7 & \xrightarrow{f} & S^7 \end{array} .$$

Then an easy verification gives

6.2. Lemma. *$b(E_f) \equiv \pm \deg f \cdot b(E)$ (mod 12).*

Now let B be a simply connected complex with $H^*(B; Z) = Z[x_4, x_8]$, $\dim x_i = i$ (e.g. $BSp(2)$). We have

$$\mathrm{KO}^*(B) = \mathrm{KO}^*(pt)\,[[x_4, x_8]] .$$

Put $u = \eta_4 x_4$, $v = \eta_8 x_4^2$, $w = \eta_8 x_8$. Using the Serre spectral sequence, we have

$$H^*(GB; Z) = \Lambda_Z[t_3, t_7] ,$$

where, under the adjoint map $SGB \to B$ and suspension, $x_4 \to t_3$, $x_8 \to t_7$. Using the same map, $u \to u$, $w \to w$ and

$$\psi^2 u = 4u + 2av + 4bw\ ,$$

where of course $\pm b$ is well determined (mod 12).

Suppose B contains an FD-subgroup $\widetilde{B}S^3$, where $\widetilde{B}S^3$ is some FD-group structure on S^3. Then we have a quotient fibration

$$S^3 \to GB \to S^7\ .$$

Now GB is an H-space; so, by a theorem of Curtis and Mislin [15], GB has the homotopy type of a principal S^3 bundle over S^7. All such are obtained as follows. Let $S^3 \to \mathrm{Sp}(2) \to S^7$ be the standard quotient of Lie groups. Let $f: S^7 \to S^7$ be of degree d and let E_d be the pull back bundle

$$\begin{array}{ccc} S^3 & = & S^3 \\ \downarrow & & \downarrow \\ E_d & \to & \mathrm{Sp}(2) \\ \downarrow & & \downarrow \\ S^7 & \to & S^7 \end{array}\ .$$

Now principal bundles over S^7 are classified by an element of $\pi_6 S^3 = Z_{12}$ with generator ω corresponding to $\mathrm{Sp}(2) \to S^7$. So E_d corresponds to $d\omega$. Using Lemma 6.7 below, $b(\mathrm{Sp}(2)) = \pm 1$ so

$$b(E_d) \equiv \pm d \pmod{12}\ .$$

The number b thus classifies the homotopy type of the space E_d since $E_{-d} \simeq E_d$. We have proved

6.3. Proposition. *If B is an FD-group structure containing a group structure on S^3, and if $H^*(B; Z) = Z[x_4, x_8]$, then GB has the homotopy type of E_d where $d \equiv \pm b(B) \pmod{12}$.*

The Hilton-Roitberg criminal is $E_{\pm 5}$. We may now prove

6.4. Theorem. *Let $\widetilde{B}S^3$ be one of the group structures on S^3 constructed in §5. Then there is a group structure on* Sp(2) *and a group structure on $E_{\pm 5}$ containing $\widetilde{B}S^3$ as an FD-subgroup structure.*

Proof. Let BSp(2) be the standard (Lie) structure on Sp(2). We have an inclusion $HP^\infty \xrightarrow{\iota} \mathrm{BSp}(2)$. $H^*(\mathrm{BSp}(2); Z) = Z[x_4, x_8]$, $H^*(HP^\infty; Z) = Z[\iota^* x_4]$, and $\iota^* x_8 = 0$. As in §5 we may construct the following commutative diagram

$$\begin{tikzcd}
 & HP^\infty \arrow[rr, dashed, "f_0"] \arrow[dl, "\iota"'] & & M_0 \arrow[rr, "f_0'"] \arrow[dl] & & K(Z,4) \arrow[dl, "\mathrm{incl}"'] \\
BSp(2) \arrow[rr, dashed, "g_0"] & & N_0 \arrow[rr] & & K(Z,4)\times K(Z,8) & \\
 & & & \widetilde{B}S^3 \arrow[uu, "g_0''"] \arrow[rr] \arrow[dl] & & M_1 \arrow[uu, "f_1''"'] \arrow[dl] \\
 & & B \arrow[uu] \arrow[rr] & & N_1 \arrow[uu, "g_1''"] & \\
 & & & & & HP^\infty \arrow[uu, "f_1"'] \arrow[dl, "\iota"] \\
 & & & & BSp(2) \arrow[uu, dashed, "g_1"] &
\end{tikzcd}$$

where

(i) f_i, f_i'', n_i are as in §5,

(ii) g_i, g_i'' represents $n_i x_4 \oplus m_i x_8$, $i = 0, 1$,

(iii) m_0 prime to 2, m_1 prime to 3, 5, 7, ...,

(iv) g_0, g_1'' are 2 equivalences,

(v) g_1, g_0'' are 3, 5, 7, ... equivalences.

By Proposition 6.3, B is a group structure on E_d, $d = b(B)$, and it is easy to see that $\widetilde{B}S^3$ is a FD-subgroup of B. It remains to calculate $b(B)$. But using Lemma 6.7 below and the technique of §5,

6.6. Lemma. $\pm b(B) \equiv 4m_1n_1 - 3m_0n_0$ *(mod 12).*

The theorem now follows from the table

$a(BS^3)$	n_0 (mod 4)	n_1 (mod 3)	m_0 (mod 4)	m_1 (mod 3)	b (mod 12)
± 1	1	1	1	1	± 1
± 1	1	1	1	-1	± 5
± 5	-1	1	-1	1	± 1
± 5	-1	1	1	1	± 5
± 7	1	-1	1	-1	± 1
± 7	1	-1	1	1	± 5
± 11	-1	-1	1	1	± 1
± 11	-1	-1	1	-1	± 5

6.5. Remark. It is not yet known whether there is a group structure on $E_{\pm 5}$ with a subtorus of maximal rank. One can, however, get an exotic "quaternionic torus" $\widetilde{B}S^3 \times \widetilde{B}S^3$ contained in a $BE_{\pm 5}$. One uses the $S^3 \times S^3 \to \mathrm{Sp}(2)$ and the above mixing procedure with m_0 and m_1 equal to 1. By comparing the table following 6.6 with that following 5.3, one sees that the $\widetilde{B}S^3$ which arise are those with $a \equiv \pm 5$ or ± 7 (mod 24).

6.7. Lemma. $b(\mathrm{BSp}(2)) = -1$.

Proof. Let T^2 be the maximal torus of Sp(2). $H^*(B_{T^2}; Z) = Z[x_1, x_2]$, $\dim x_i = 2$. Let $H^*(\mathrm{BSp}(2), Z) = Z[t_4, t_8]$ and, if ι is the inclusion of T^2, $\iota^* t_4 = x_1^2 + x_2^2$, $\iota^* t_8 = x_1^2 x_2^2$. We have the complex representation ring $\mathrm{RU}(T^2) = Z[z_1, z_1^{-1}, z_2, z_2^{-1}]$, and $\mathrm{KU}(B_T) = \hat{\mathrm{RU}}(T^2)$, with $\mathrm{KU}(B_T) = Z[[\xi_1, \xi_2]]$, $\xi_1 = z_1 - 1$, $\xi_2 = z_2 - 1$, where ξ_i represents x_i. Now [2, 7.6] $\mathrm{RU}(\mathrm{Sp}(2)) = Z[\lambda_1, \lambda_2]$, where λ_i = are the ith symmetric functions of $z_1, z_1^{-1}, z_2, z_2^{-1}$. Furthermore, λ_1 is symplectic and λ_2 real. One calculates that

$$\eta_1 = (\lambda_1 - 4)$$
$$\eta_2 = (\lambda_2 - 2\lambda_1 + 2)$$

represent $\iota^* t_4 = x_1^2 + x_2^2$ and $\iota^* t_8 = x_1^2 x_2^2$ respectively. Furthermore, η_1 is symplectic and η_2 real. Calculate

$$\psi^2 \eta_1 = 4\eta_1 + \eta_1^2 - 2\eta_2 .$$

Transferring this information to KO-theory via the results of [4] gives the lemma.

Appendix

Fibre homotopy equivalence of simplicial fibrations

We wish to have in simplicial language the results of Dold and Lashoff [16]. Let X be a Kan complex. We have a complex X^X of maps of X into X. An n-simplex of X^X is a fibre map

$$\begin{array}{ccc} X \times \Delta_n & \xrightarrow{f} & X \times \Delta_n \\ \downarrow & & \downarrow \\ \Delta_n & \longrightarrow & \Delta_n \end{array}$$

where Δ_n is the standard n-simplex. A theorem of Moore [e.g. 24, p.69] states the X^X is a Kan-complex.

A.1. Definition. The complex $H(X)$ of homotopy equivalences of X is the subcomplex of X^X of maps f above such that f restricts to a homotopy equivalence on the fibre over a vertex of Δ_n. Clearly

A.2. Lemma. *$H(X)$ is a simplicial monoid and a Kan complex.*

Moore [26] gives a construction of a principal classifying fibration

$$H(X) \to WH(X) \to \bar{W}H(X) = B_{H(X)}$$

so we may talk about bundles with fibre X and structure monoid $H(X)$. We wish to prove

A.3. Proposition. *Fibre homotopy classes of Kan fibrations*

$$X \to E \to B$$

with fibre X and base B are in one to one correspondence with homotopy classes of maps $B \to B_{H(X)}$.

Proof. We must show that every fibration is fibre homotopy equivalent to an $H(X)$ bundle and any $H(X)$ bundles which are fibre homotopy equivalent are equivalent. Let F,

$$F \xrightarrow{i} X \xrightarrow{j} F$$

be a minimal subcomplex, j a retraction, $ji = 1$, $ij \sim 1$. Then i and j induce a homomorphism

$$H(F) \to H(X)\,,$$

$$\alpha \to i \circ \alpha \circ j\,.$$

Since $i \circ j \sim 1$, this map is a homotopy equivalence. Therefore we have a homotopy equivalence

$$B_{H(F)} \to B_{H(X)}\,.$$

But clearly $H(F) = A(F)$ [8], the automorphism complex of F, since F is minimal. But $B_{A(F)}$ classifies isomorphism classes of minimal fibrations [8], and fibre homotopy equivalences of minimal fibrations are isomorphisms. Furthermore, every Kan fibration has a minimal fibration as a deformation retraction [26]. The proposition now follows.

One would like to relate this to the topological case. Let RX be the geometric realization of X. For any n-simplex $f\colon X \times \Delta_n \to X \times \Delta_n$ of X^X we can define a projection $\tilde{f}\colon X \times \Delta_n \to X$ and thus a map

$$R\tilde{f}\colon RX \times R\Delta_n \to RX\,.$$

But $R\Delta_n$ is the standard topological n-simplex so we have defined a singular simplex $R\widetilde{f} \in \mathrm{Sin}\, RX^{RX}$, where Sin denotes the singular complex. We thus have a simplicial map $H(X) \to \mathrm{Sin}\, H(RX)$, where $H(RX)$ is the space of homotopy equivalences of RX. There is an adjoint map

$$RH(X) \to H(RX)\,.$$

A folk conjecture is

A.3. Proposition. *The above map*

$$RH(X) \to H(RX)$$

is a weak homotopy equivalence.

The author intends to prove this in a future paper.

References

[1] J.F. Adams, On complex Stiefel manifolds, Proc. Cambridge Phil. Soc. 61 (1965) 81–103.
[2] J.F. Adams, Lectures on Lie groups (Benjamin, New York, 1969).
[3] J.F. Adams, Vector fields on spheres, Ann. of Math. 75 (1962) 604–632.
[4] D.W. Anderson, The real K-theory of classifying spaces, Proc. Nat. Acad. Sci. USA 51 (1964) 634–636.
[5] E. Artin, Galois theory, Notes, University of Notre Dame, 1959.
[6] M.F. Atiyah, K-theory (Benjamin, New York, 1967).
[7] M.F. Atiyah and F. Hirzebruch, Vector bundles and homogeneous spaces, Proc. Symp. in Pure Mathematics 3, Differential geometry (Amer. Math. Soc., 1961) pp. 7–38.
[8] M. Barratt, V. Gugenheim and J.C. Moore, On semisimplicial Fibre bundles, Amer. J. Math. 81 (1959) 639–657.
[9] P. Baum, Cohomology of homogeneous spaces, Bull. Amer. Math. Soc. 69 (1963) 531–533.
[10] P. Baum, On the cohomology of homogeneous spaces, Topology 7 (1968) 15–38.
[11] A. Borel, Sur la cohomologie des espaces fibre principaux et des espaces homogènes de groups de Lie compacts, Ann. of Math. 57 (1953) 115–207.
[12] A. Borel, Topics in the homology theory of fibre bundles, Lecture notes in Mathematics, Vol. 36 (Springer-Verlag, Berlin, 1967).
[13] R. Bott, Lectures on $K(X)$ (Benjamin, New York, 1969).
[14] W. Browder, Torsion in H-spaces, Ann. of Math. 74 (1961) 24–51.
[15] M.L. Curtis and G. Mislin, H-spaces which are bundles over S^7, Preprint, Rice University 1969.
[16] A. Dold and R. Lashof, Principal quasifibrations and fibre homotopy equivalences of bundles, Illinois J. Math. 3 (1959) 285–305.
[17] A. Dold and R. Thom, Quasifaserungen und unendliche symmetrische Produkte, Ann. of Math. 67 (1958) 239–281.
[18] P. Hilton and J. Roitberg, On principle S^3-bundles over spheres, Ann. of Math. 90 (1969) 91–107.
[19] H. Hopf, Über die Topologie der gruppen-Mannigfaltigkeiten und ihren Verallgemeinerungen, Ann of Math. 42 (1941) 22–52.
[20] J. Hubbuck, On homotopy commutative H-spaces, Topology 8 (1969) 119–126.

[21] D.M. Kan, A combinatorial definition of homotopy groups, Ann. of Math. 67 (1958) 282–312.

[22] D.M. Kan, On homotopy theory and c.s.s. groups, Ann. of Math. 68 (1958) 38–53.

[23] V.J. Lal, Wall obstruction of a fibration, Inventiones Math. 6 (1968) 67–77.

[24] J.P. May, Simplicial objects in algebraic topology (Van Nostrand, Princeton, N.J., 1967).

[25] J. Milnor, Construction of universal bundles II, Ann. of Math. 63 (1956) 430–436.

[26] J.C. Moore, Semi-simplicial complexes and postnikov systems, Symp. Intern. de Topologia algebraica, Universidad Nacional Autonoma de Mexico, Mexico City, 1958, pp. 232–247.

[27] J.P. Serre, Groups d'homotopie et classes de groups Abéliens, Ann. of Math. 58 (1953) 258–294.

[28] J. Slifker, Exotic multiplications on S^3, Quart. J. Math. Oxford (2), 16 (1965) 322–359.

[29] J. Stasheff, Homotopy associativity of H-spaces, I, II, Trans. Amer. Math. Soc. 108 (1963) 275–312.

[30] J. Stasheff, Manifolds of the homotopy type of (non-Lie) groups, Bull. Amer. Math. Soc. 75 (1969) 998–1000.

[31] R. Stong, Notes on cobordism theory (Princeton University Press, 1968).

[32] B.L. van der Waerden, Modern algebra, Vol. 2, 2nd ed. (Ungar, New York, 1950).

[33] C.T.C. Wall, Finiteness conditions for CW-complexes II, Proc. Royal Soc. London 295 (1966) 129–139.

[34] A. Zabrodsky, Homotopy associativity and finite CW complexes, Topology 9 (1970) 121–128.

[35] E. Zeeman, A proof of the comparison theorem for spectral sequences, Proc. Cambridge Phil. Soc. 53 (1957) 57–62.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 3 (1971) pp. 275–279.

EMBEDDING NONSINGULAR MODULES IN FREE MODULES *

K.R. GOODEARL
University of Washington, Seattle, Wash. 98105, USA

Received 22 October 1970

Introduction

It is easily shown that any finitely generated torsion-free module over a commutative integral domain may be embedded in a free module. A well-known extension of the concept of a torsion-free module is the idea of a nonsingular module [5,6]. We consider the above result in this context, where it does not always hold, and prove the following characterization:

If R is a ring with zero right singular ideal, then every finitely generated nonsingular right R-module can be embedded in a free right R-module iff Q_R is flat and $(Q \otimes_R Q)_R$ is nonsingular, where Q is the maximal right quotient ring of R.

Preliminaries

R will be an associative ring with unit, and all modules will be unitary. All tensor products will be taken over R. We shall use $E(A)$ to denote the injective hull of a module A. Q will denote the maximal right quotient ring of R.

Let S $\{S'\}$ denote the set of all essential right $\{$left$\}$ ideals of R. Recall that the *singular submodule* of any right $\{$left$\}$ R-module A is the set $Z(A)$ $\{Z'(A)\}$ of those elements of A which are annihilated by some member of S $\{S'\}$. A is said to be *singular* $\{$*nonsingular*$\}$ if its singular submodule is A $\{0\}$.

N.B.: We shall assume throughout this paper that R_R is nonsingular. However, we do not assume that ${}_RR$ is nonsingular, but will prove this in certain circumstances.

Remark (a). If A is an essential submodule of B, then B/A is singular. If A is also non-singular, then so is B.

Theorem 1. *Q is a regular ring, Q_Q is injective, and $Q_R = E(R_R)$.*

Proof. [4], Theorem 1 + 2, p. 69.

* This research was supported by an N.S.F. Graduate Fellowship, and constitutes a portion of the author's doctoral dissertation.

Proposition 2. (a) *A_R is singular iff* $\mathrm{Hom}_R(A, Q) = 0$. (b) *Any extension of singular modules is singular.* (c) *For any A_R, $Z(A/Z(A)) = 0$.*

Proof. (a) follows from Proposition 1 of [5], and then (b) and (c) follow easily.

Theorem 3. *If A is a finitely generated right Q-module such that $Z(A) = 0$, then A_Q is projective and injective.*

Proof. One checks that A is nonsingular as a Q-module and then applies [8], Theorem 2.7.

Theorem 4. *If A_R is finitely generated and nonsingular, then A can be embedded in a finite direct sum of copies of Q_R.*

Proof. [2], Lemma 2.2.

Lemma 5. *If $J \leqslant I \in S$ and $(I/J) \otimes Q = 0$, then $J \in S$.*

Proof. By [7], Proposition 2.2, I/J is singular, hence Proposition 2 says R/J is singular.

Lemma 6. *Let B be a right Q-module with $Z(B) = 0$. If A is any R-submodule of B, then $Z(A \otimes Q)$ is the kernel of the natural map $A \otimes Q \to B$.*

Proof. This is proved in exactly the same manner as [1], Lemma 1.8.

Remark (b). Lemma 6 says that $Z(Q \otimes Q) = 0$ iff the multiplication map $Q \otimes Q \to Q$ is an isomorphism.

The main theorem

Theorem 7. *Let R be a ring with $Z(R_R) = 0$, Q its maximal right quotient ring. Then, every finitely generated nonsingular right R-module can be embedded in a free right R-module iff Q_R is flat and $Z[(Q \otimes_R Q)_R] = 0$.*

Proof. Necessity: Let $j\colon {}_RI \to {}_RR$ be an inclusion, $x \in Q \otimes I$ such that $(1 \otimes j)(x) = 0$. $\exists$ a finitely generated $A_R \leqslant Q$ (with inclusion $k\colon A \to Q$) and $y \in A \otimes I$ with $(k \otimes 1)(y) = x$. $\exists$ a monomorphism $f\colon A \to F$ with F_R free. Extend $f^{-1}\colon fA \to A$ to $g\colon F \to Q$, and note that $gf = k$. Then we have a commutative diagram (diagram 1) with h and i monic. A simple diagram chase yields $x = 0$.

$$
\begin{tikzcd}
A \otimes I \arrow[rr, "k \otimes 1"] \arrow[dr, "f \otimes 1"] \arrow[ddd, "1 \otimes j"'] & & Q \otimes I \arrow[ddd, "1 \otimes j"] \\
& F \otimes I \arrow[ur, "g \otimes 1"'] \arrow[d, "h = 1 \otimes j"] & \\
& F \otimes R \arrow[dr, "g \otimes 1"] & \\
A \otimes R \arrow[ur, "f \otimes 1"] \arrow[rr, "i = k \otimes 1"'] & & Q \otimes R
\end{tikzcd}
$$

Diagram 1.

Next consider any $x \in Z(Q \otimes Q)$. Find A, k, y, f, g as above (replacing I by Q). Then we have a commutative diagram (diagram 2) in which the columns are exact by Lemma 6. $Z(F \otimes Q) = 0$ since $(F \otimes Q)_Q$ is free, and then another simple diagram chase yields $x = 0$.

$$\begin{tikzcd}
Z(A \otimes Q) \arrow[d] & F \otimes Q \arrow[dr, "g \otimes 1"] & Z(Q \otimes Q) \arrow[d] \\
A \otimes Q \arrow[ur, "f \otimes 1"] \arrow[rr, "k \otimes 1"] \arrow[d] & & Q \otimes Q \arrow[d] \\
Q \arrow[rr, "1"] & & Q
\end{tikzcd}$$

Diagram 2.

The proof of sufficiency requires several intermediate results. For Lemmas 8–11, we make the standing hypothesis that Q_R is flat and $Z(Q \otimes Q) = 0$. P will denote the maximal left quotient ring of Q.

Lemma 8. *Q is a left quotient ring of R.*

Proof. [3], Proposition 2.2.

Lemma 9. *$Z'(R) = Z'(Q) = Z'(P) = 0$, and P is the maximal left quotient ring of R. For any $a_1, \ldots, a_n \in P$, $I = \{r \in R \mid ra_i \in R \forall i\} \in S'$.*

Proof. $Z'(R) = 0$ by [4], Proposition 3, p. 70. Then $Z'(Q) = 0$ by remark (a). P is the maximal left quotient ring of R by [4], Proposition D, p. 66. Now ${}_RR$ is essential in ${}_RP$, so remark (a) says that $Z'(P) = 0$ and ${}_R(P/R)$ is singular. Then each left ideal $\{r \in R \mid ra_i \in R\}$ is in S', hence so is their intersection, i.e., I.

Lemma 10. *P_R is flat and $Z(P) = Z(P \otimes Q) = Z'(P \otimes Q) = 0$.*

Proof. P_Q is flat since Q is regular, so Q_R flat $\to P_R$ flat. Since ${}_RR$ is essential in ${}_RP$, any nonzero cyclic submodule of P_R has a nonzero homomorphism into R_R (induced by a left multiplication), hence $Z(P) = 0$.

Consider any $x \in Z(P \otimes Q)$. $\exists$ a finitely generated $A_Q \leqslant P$ (with inclusion $k: A \to P$) and $y \in A \otimes Q$ such that $(k \otimes 1)(y) = x$. A_Q is projective and injective by Theorem 3. Then A_Q is a summand of P_Q, hence $k \otimes 1$ is monic and $y \in Z(A \otimes Q)$. But $(A \otimes Q)_R$ is a summand of a direct sum of copies of $(Q \otimes Q)_R$, hence $Z(A \otimes Q) = 0$. Thus $x = 0$.

Lemma 6 now says the multiplication map $P \otimes Q \to P$ is an isomorphism, so $Z'(P \otimes Q) = 0$.

Lemma 11. *Given $a_1, \ldots, a_n \in Q$, $\exists$ a finitely generated $J \in S'$ such that $Ja_i \leqslant R \; \forall\, i$.*

Proof. Set $I=\{r\in R \mid ra_i\in R\ \forall i\}$, $R_i={}_RR$ for $i=0,\ldots,n$, $Q_i={}_RQ$ for $i=1,\ldots,n$, $F=R_0\oplus\ldots\oplus R_n$. Note that $I\in S'$ by Lemma 9. $\exists$ an exact sequence

$$0\to I\xrightarrow{f}F\xrightarrow{g}Q_1\oplus\ldots\oplus Q_n$$

defined by $f(r)=(r,ra_1,ra_2,\ldots,ra_n)$ and $g(r_0,r_1,\ldots,r_n)=(r_0a_1-r_1,r_0a_2-r_2,\ldots,r_0a_n-r_n)$. Set $A=gF$. By Lemma 10, P_R is flat and each $Z'(P\otimes Q_i)=0$, so

$$P\otimes A\to(P\otimes Q_1)\oplus\ldots\oplus(P\otimes Q_n)$$

is monic and $Z'(P\otimes A)=0$.

Set $P_i={}_RP$ for $i=0,\ldots,n$. $\exists$ an exact sequence

$$0\to P\xrightarrow{h}P_0\oplus\ldots\oplus P_n\xrightarrow{k}P_1\oplus\ldots\oplus P_n$$

defined by $h(s)=(s,sa_1,sa_2,\ldots,sa_n)$ and $k(s_0,s_1,\ldots,s_n)=(s_0a_1-s_1,s_0a_2-s_2,\ldots,s_0a_n-s_n)$. We have a commutative diagram (diagram 3) with exact rows. j is monic by Lemma 6, so i is an isomorphism and ${}_R(P\otimes I)$ is cyclic. Then $\exists$ a finitely generated $J\leqslant I$ such that $P\otimes(I/J)=0$. By Lemma 5, $J\in S'$.

$$\begin{tikzcd}
0 \arrow[r] & P\otimes I \arrow[r] \arrow[d, "i"] & P\otimes F \arrow[r] \arrow[d, "\cong"] & P\otimes A \arrow[r] \arrow[d, "j"] & 0 \\
0 \arrow[r] & P \arrow[r] & P_0\oplus\ldots\oplus P_n \arrow[r] & P_1\oplus\ldots\oplus P_n &
\end{tikzcd}$$

Diagram 3.

We are now in a position to complete the proof of Theorem 7. Thus let A_R be finitely generated and nonsingular. By Theorem 4, it suffices to consider the case $A\leqslant Q$. By Lemma 11 $\exists$ a finitely generated $J\in S'$ such that $JA\leqslant R$. If $b_1,\ldots,b_m$ generate ${}_RJ$, then each b_i induces a homomorphism $f_i: A\to R_R$. The map $\oplus f_i: A\to\oplus R_R$ is monic since $Z'(Q)=0$.

Remark (c). V.C. Cateforis ([2], Theorem 1.6) proves several conditions equivalent to "${}_RQ$ is flat and $Z(Q\otimes_R Q)=0$". In case R is commutative, these conditions will then be equivalent to the assertions of Theorem 7.

Corollary 12. Let R be a ring for which R_R is finite-dimensional and nonsingular. Then, every finitely generated nonsingular right R-module can be embedded in a free right R-module iff $E(R_R)$ is flat.

Proof. Q is a semisimple ring by [7], Theorem 1.6. Thus $(Q\otimes Q)_Q$ is projective, so $Z(Q\otimes Q)=0$.

The following result of V.C. Cateforis ([2], Theorem 2.3) is an immediate consequence of Theorem 7:

Theorem 13. *For a ring R with maximal right quotient ring Q, the following are equivalent:*

(a) $Z(R_R) = 0$ *and every finitely generated nonsingular right R-module is projective.*

(b) *R is right semihereditary,* Q_R *is flat, and* $Z[(Q \otimes_R Q)_R] = 0$.

Bibliography

[1] V.C. Cateforis, Flat regular quotient rings, Trans. Am. Math. Soc. 138 (1969) 241–249.

[2] V.C. Cateforis, On regular self-injective rings, Pacific J. Math. 30 (1969) 39–45.

[3] V.C. Cateforis, Two-sided semisimple maximal quotient rings, Trans. Am. Math. Soc. 149 (1970) 339–349.

[4] C. Faith, Lectures on injective modules and quotient rings (Springer-Verlag, New York, 1967).

[5] F.R. Gentile, Singular submodule and injective hull, Indag. Math. 24 (1962) 426–433.

[6] A.W. Goldie, Torsion-free modules and rings, J. Algebra 1 (1964) 268–287.

[7] F.L. Sandomierski, Semisimple maximal quotient rings, Trans. Am. Math. Soc. 128 (1967) 112–120.

[8] F.L. Sandomierski, Nonsingular rings, P.A.M.S. 19 (1968) 225–230.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 3 (1971) pp. 281–316.

CONVENIENT CATEGORIES OF TOPOLOGICAL ALGEBRAS, AND THEIR DUALITY THEORY

Eduardo J. DUBUC
Dalhousie University, Halifax, N.S., Canada

and

Horacio PORTA *
University of Illinois, Urbana, Ill., USA

Received 18 January 1971

Introduction

Concrete algebraic structures with a topology have long arisen in mathematical practice, leading to the notion of a topological space with algebraic operations making the underlying set an algebra for the type under consideration. Classes of such objects (together with continuous maps respecting the algebraic structure) form categories which, understandably, do not share some important properties of their purely algebraic analogues. Specially, *their relation with the base category* ***S*** *of sets is not satisfactory* (e.g., they are not monadic (i.e., tripleable) with respect to the natural forgetful functors). This is essentially due to the fact that taking forgetful functors into ***S*** is forgetting too much. Of importance is also the fact that the set of morphisms between any two such algebras carries a topology which is inherited from the topologies of the algebras, and which is not taken into account (it is ignored). That is, the ubiquitous "always at our disposal, no need to be defined" representable functors do not retain any topological information.

The category of topological spaces is actually the natural *base* category (that is, the place where the forgetful and representable functors land) for a categorical approach to the study of classes of topologized algebraic structures. However, this category is not "set-like" enough to make such an approach possible. Categories which, like ***S***, have enough structure to serve as base categories have been recognized by category theorists during the sixties, when the concept of *closed category* was developed. The category ***K*** of compactly generated Hausdorff topological spaces is such a convenient (closed) category.

* Research supported by a National Science Foundation grant (NSF GP 20431).

The study of enriched category theory has reached a level of development which puts at our disposal enriched versions of most of the important machinery of ordinary category theory. If the base category V is good enough, the V-world is as good as the set-based world. Although not completely, this is very much the case with K. *Working in the K-world allows us to deal with topologized algebraic structures in a purely algebraic way.* The continuity of the functions is always guaranteed, the topology in the constructions does not require ad-hoc definition. The topological information is carried automatically due to the closed structure of K. The (left) adjoints to the representable functors, currently called *tensors* and *cotensors*, can not (in the K-world) be obtained as colimits and limits. Thus, they provide a (categorical) characterization for certain constructions which is not available (or even possible) in an ordinary set based approach. The algebraically defined categories of groups in K, modules in K over a ring in K, algebras in K over a ring in K, etc. etc., are all K-monadic (i.e., K-tripleable) with respect to the natural "forget-the-algebraic-structure" forgetful K-functors into K, in exactly the same way that the analogous categories in the set based world.

We introduce here a systematic treatment of categories of (complex) topological algebras considered as categories based in the category K of compactly generated Hausdorff spaces. This leads to the definition of K-topological algebras (i.e., the concept of associative algebra over the field of complex numbers ***relativized*** to the K-world). Roughly, a K-topological algebra turns out to be a complex algebra with a topology making the operations continuous when restricted to compact sets. This is a broad class of algebras, containing *all* algebras with jointly continuous product, and also many interesting algebras with discontinuous, separately continuous, product (cf. Examples 1.3 and 1.4).

Cotensors realize the (classical) construction of algebras of continuous functions, and play an important role in duality theory, as illustrated by the following simple formulation of the main result in Gelfand Theory: "The complex numbers are a K-codense K-cogenerator of the K-category of commutative C^*-algebras with identity".

In Section 1 we introduce our basic definitions and show how K-topological algebras (and subclasses like algebras with involution, Banach algebras, Fréchet algebras, C^*-algebras, etc.) form K-categories, in Section 2 we establish some properties of these K-categories, and in Section 3 we show that any complete locally m-convex commutative algebra with identity, and with an involution such that for a defining family of seminorms $\{p\}$ the identity $p(xx^*) = p(x)^2$ holds, is the algebra of *all* continuous complex valued functions, with the uniform convergence on compact sets, on a certain topological space (topologically as well as algebraically, if considered qua K-topological algebras) (cf. Theorem 3.13). This is done by interpreting functional representation within the general framework of an (enriched) duality machinery. This machinery is basically the interplay of the cotensors with the contravariant representable functor determined by the complex numbers (which realize the classical construction of the spectrum of an algebra), in what could be called an iterated double dualization process.

Let us remark that the words "topology", "topologized", "continuous", etc., could have been omitted (except in the examples, of course) immediately after Section 0, where we collect some known basic properties of $\boldsymbol{K}$. We did not do so in order to remind the reader (and ourselves) that we are dealing with topological spaces and continuous functions.

Finally, we point out that other categorical approaches to standard Functional Analysis theories have been exploited, for instance, in the recent papers [25, 29, 30].

§0. The category K

We will denote by $\boldsymbol{K}$ the full subcategory of topological spaces whose objects are the *compactly generated Hausdorff* spaces, that is: $X \in \boldsymbol{K}$, if and only if X is a Hausdorff space and $X = \underset{K \subset X}{\underrightarrow{\text{colim}}} K$, K running over all the compact subsets of X. This means that X is topologically equal to the topological colimit of its compact subsets; in other words, a map $X \to Z$ into any other topological space Z is continuous if and only if it is continuous on each compact subset of X. Following [16], we will call such a space Kelley space. Given any Hausdorff space Z is clear that Z, ***as a set*** is equal to the colimit of its compact subsets. The colimit topology defines a Kelley space denoted KeZ, and the inclusions $K \to Z$ of the compact subsets of Z determine a (unique, continuous) map Ke$Z = \underset{K \subset Z}{\underrightarrow{\text{colim}}} K \to Z$. Z is a Kelley space if and only if Z = KeZ. KeZ has the same underlying set as Z, and its topology is the finest among those having the same compact subsets as the given topology of Z. Given any topological space H, a map Ke$Z \to H$ is continuous if and only if $Z \to H$ is continuous on compact subsets. Also: given any Kelley space X, a map $X \to Z$ is continuous if and only if $X \to$ KeZ is continuous. Denoting by $\mathbf{Top}_2$ the category of Hausdorff spaces and continuous maps, for any $Z \in \mathbf{Top}_2$, the assignment $Z \rightsquigarrow$ KeZ is then a functor $\mathbf{Top}_2 \xrightarrow{\text{Ke}} \boldsymbol{K}$ (Ke$f = f$, for any $Z \xrightarrow{f} Z'$) which provides a right adjoint (coreflexion) to the full inclusion $\boldsymbol{K} \to \mathbf{Top}_2$. We call this functor the *K-ation functor*, and for $Z \in \mathbf{Top}_2$, KeZ will be the *K-ation of Z*

We will give below a list of properties of the category $\boldsymbol{K}$. The reader is referred to [16] for quick proofs of the results below. For a more extensive treatment and additional results, he can use [31], [35], [37], and for a treatment with a categorical flavor, [32] and [33].

The categorical language and terminology used here is by now standard in articles written in English on this side of the Atlantic. The basic categorical concepts can be found in [26]. For the notion of closed category and related subjects there is a condensed presentation in [10]. A more extensive presentation is given in [6]. On the other hand, [14] is a complete, exhaustive and meticulous reference article. [7], [10] and [23] consider further developments of the subject.

If X and Y are Kelley spaces, the set of all continuous functions from X to Y will be denoted by $\boldsymbol{K}_0(X, Y)$. Thus, $\boldsymbol{K}_0(X, Y) = \mathbf{Top}_2(X, Y)$.

0.0. *If $X \in \boldsymbol{K}$, then $C \subset Z$ is closed if and only if $C \cap K$ is closed in X for all K compact in X. For $Z \in$* **Top**$_2$, *the family $\{C \subset Z; C \cap K$ is closed in Z for all K compact in $Z\}$ is a basis of closed sets for* KeZ.

0.1. *Metrizable spaces are Kelley spaces.*

0.2. *Locally compact Hausdorff spaces are Kelley spaces.*

0.3. *Any Hausdorff quotient of a Kelley space is a Kelley space.*

0.4. *A closed subset of a Kelley space is a Kelley space with the induced topology.*

0.5. *An open subset of a Kelley space is a Kelley space with the induced topology.*

0.6. Definition. *Given a continuous monomorphism $X \xrightarrow{i} Y$ between Kelley spaces, i is a full injection if given any other Kelley space V, a function $V \to X$ is continuous if and only if the composite $V \to X \xrightarrow{i} Y$ is continuous.* This is equivalent to saying that the topology of X is the K-ation of the inverse image under i of the topology of Y. A topological subspace is a full injection, but the converse does not hold.

0.7. *$\boldsymbol{K}$ is a complete category.* That is: $\boldsymbol{K}$ has all (small) inverse limits (limits). If $\Lambda \to \boldsymbol{K}$ ($\lambda \rightsquigarrow X_\lambda$) is a functor, then $\varprojlim_\lambda X_\lambda \in \boldsymbol{K}$ is the limit space of the X_λ with the K-ation of the limit topology (which is automatically Hausdorff).

0.8. *$\boldsymbol{K}$ is a cocomplete category.* That is: $\boldsymbol{K}$ has all (small) direct limits (colimits). If $\Lambda \to \boldsymbol{K}$ ($\lambda \rightsquigarrow X_\lambda$) is a functor, $\operatorname{colim}_{\vec{\lambda}} X_\lambda \in \boldsymbol{K}$ is the largest Hausdorff quotient of the colimit space of the X_λ with the colimit topology (which is automatically compactly-generated).

0.9. *Given two Kelley spaces X and Y we will denote by $X \boxtimes Y$ the product of X and Y in $\boldsymbol{K}$ (the existence of which follows from 0.7).* We have $X \boxtimes Y = \mathrm{Ke}(X \times Y)$, where $X \times Y$ denotes the ordinary cartesian product. If $\tau: X \boxtimes Y \to Y \boxtimes X$ is defined by $\tau(x, y) = (y, x)$, then τ is an isomorphism, or $X \boxtimes Y \overset{\tau}{\approx} Y \boxtimes X$. The Kelley space consisting of a single point $\{*\}$ will be denoted by 1. It is a *terminal object* of $\boldsymbol{K}$ and $X \boxtimes 1 \approx 1 \boxtimes X \approx X$ for all $X \in \boldsymbol{K}$.

0.10. *Given two Kelley spaces X and Y, $X \boxtimes Y = \operatorname{colim}_{\overrightarrow{K,K'}} K \times K'$ where $K \subset X$, $K' \subset Y$ range over all the compact subsets of X and Y.*

0.11. *If X is locally compact (whence $X \in \boldsymbol{K}$, see 0.2), then $X \boxtimes Y = X \times Y$ for all $Y \in \boldsymbol{K}$* and in fact this property characterizes locally compact spaces.

0.12. *For all* $X\in K$, *the functor* $K \xrightarrow{-\boxtimes X} K$, $Y \rightsquigarrow Y\boxtimes X$ *has a right adjoint* $K \xrightarrow{K(X,\ )} K$, $V \rightsquigarrow K(X,V)$ *where* $K(X,V)$ *is the space of all continuous functions* $X\to V$ *with the K-ation of the compact-open topology.* Thus $K_0(Y\boxtimes X, V) \overset{\omega_0}{\approx} K_0(Y, K(X,V))$ with $\omega_0\circ\omega_0 = \mathrm{id}$ (we denote this bijection with the same letter in both directions). We will also write:

$$\frac{Y\boxtimes X\to V}{Y\to K(X,V)}\ \omega_0 .$$

0.13. It follows (categorically) that the *above bijection is actually a (natural) homeomorphism* $K(Y\boxtimes X, V)\overset{\omega}{\approx} K(Y, K(X,V))$. *There are also homeomorphisms* $K(Y, K(X,V)) \overset{\alpha}{\approx} K(X, K(Y,V))$ *and the (obviously defined) maps in the following list are well defined and continuous:*

$$Y\to K(K(X,V),V)$$

$$Y\to K(X, Y\boxtimes X)$$

$$K(X,Y)\boxtimes X\to Y$$

$$K(X,Y)\boxtimes K(Y,V)\to K(X,V)$$

etc.

0.14. Proposition. *If* $X\xrightarrow{i} Y$ *is a full injection (see Def. 0.6) then* $K(V,X)\xrightarrow{K(V,i)} K(V,Y)$ *is also a full injection for all* $V\in K$.

Proof. Let $W\in K$ be any Kelley space and consider a commutative diagram

$$\begin{tikzcd} W \arrow[r, dashed, "f"] \arrow[dr, "g"'] & K(V,X) \arrow[d, "K(V,i)"] \\ & K(V,Y) \end{tikzcd}$$

where g is continuous. The proposition will follow if we prove that f is necessarily continuous. Consider now the diagram

$$\begin{tikzcd} W\boxtimes V \arrow[r, dashed, "h"] \arrow[dr, "\omega_0(g)"'] & X \arrow[d, "i"] \\ & Y \end{tikzcd}$$

where ω_0 is the bijection of 0.12 and h is the function $(x,v)\rightsquigarrow f(x)(v)$. The diagram clearly commutes, and hence h is continuous. By naturality of ω_0 the

diagram

$$\begin{tikzcd} W \arrow[r, "\omega_0(h)"] \arrow[dr, "g"'] & K(V,X) \arrow[d, "K(V,f)"] \\ & K(V,Y) \end{tikzcd}$$

also commutes. It follows that $f = \omega_0(h)$ and therefore f is continuous, as desired. Q.E.D.

We summarize: the category $\boldsymbol{K}$ is a *symmetrical monoidal closed category* (0.9 and 0.12) with a tensor product given by the (categorical) product (0.9). That is: $\boldsymbol{K}$ is a *cartesian closed category*. Furthermore, $\boldsymbol{K}$ is a *complete and cocomplete category* (0.7 and 0.8'). Finally, it is easy to observe that $1 \in \boldsymbol{K}$ is a *generator* and that $\boldsymbol{K}$ is *well-powered* (i.e., the class of subobjects of any fixed object of $\boldsymbol{K}$ is a set).

Observe that $\boldsymbol{K}$ is equivalent to the category of *all* Hausdorff spaces and all functions that are continuous on compact subsets, between them. More precisely, denoting by $\boldsymbol{K}\mathbf{Top}_2$ this category, the inclusion $\boldsymbol{K} \to \boldsymbol{K}\mathbf{Top}_2$ is still full and the K-ation functor Ke is also a functor $\boldsymbol{K}\mathbf{Top}_2 \xrightarrow{\mathrm{Ke}} \boldsymbol{K}$. The map $\mathrm{Ke}Z \to Z$ (for $Z \in \boldsymbol{K}\mathbf{Top}_2$) is an isomorphism in $\boldsymbol{K}\mathbf{Top}_2$ and therefore Ke (together with the inclusion) is an equivalence of categories. Thus, *the choice between $\boldsymbol{K}$ and $\boldsymbol{K}\mathbf{Top}_2$ is just a matter of personal taste.*

§ 1. Categories of topological algebras

The field of complex numbers with the ordinary (metric) topology is a Kelley space, that we will denote by C.

1.1. Definition. *By a K-topological algebra we will understand an algebra over C in $\boldsymbol{K}$. Specifically, a K-topological algebra consists of the following:*

A Kelley space X (that is, an object of $\boldsymbol{K}$) together with

1.1.1. *maps (in $\boldsymbol{K}$)* $X \boxtimes X \xrightarrow{+} X$

$1 \xrightarrow{0} X$

$X \xrightarrow{-} X$

such that

(i) *+ is associative and commutative. That is, the diagrams*

$$\begin{tikzcd} X \boxtimes X \boxtimes X \arrow[r, "\mathrm{id} \boxtimes +"] \arrow[d, "+ \boxtimes \mathrm{id}"'] & X \boxtimes X \arrow[d, "+"] \\ X \boxtimes X \arrow[r, "+"] & X \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} X \boxtimes X \arrow[r, "\tau"] \arrow[dr, "+"'] & X \boxtimes X \arrow[d, "+"] \\ & X \end{tikzcd}$$

commute (where τ is defined as in 0.9).

(ii) *0 is a unit for $+$. That is, the diagram*

$$\begin{tikzcd}
X \boxtimes X \arrow[dr, "+"'] & X \boxtimes 1 = X = 1 \boxtimes X \arrow[l, "\mathrm{id} \boxtimes 0"'] \arrow[r, "0 \boxtimes \mathrm{id}"] \arrow[d, "\mathrm{id}"] & X \boxtimes X \arrow[dl, "+"] \\
& X &
\end{tikzcd}$$

commutes.

(iii) *$-$ is an inverse for $+$ with respect to 0. That is, the diagram*

$$\begin{tikzcd}
X \boxtimes X \arrow[drr, "+"'] & X \boxtimes X \arrow[l, "- \boxtimes \mathrm{id}"'] & X \arrow[l, "\Delta"'] \arrow[r, "\Delta"] \arrow[d, "0"] & X \boxtimes X \arrow[r, "\mathrm{id} \boxtimes -"] & X \boxtimes X \arrow[dll, "+"] \\
& & X & &
\end{tikzcd}$$

commutes (where $\Delta(x) = (x, x)$ and $X \xrightarrow{0} X = X \to 1 \xrightarrow{0} X$).

1.1.2. *a map (in K) $X \boxtimes X \xrightarrow{\cdot} X$ such that*

(i) *$\cdot$ is associative (in the same sense as $+$; see 1.1.1.(i)).*

(ii) *$\cdot$ is distributive with respect to $+$. That is, the diagram*

$$\begin{tikzcd}
X \boxtimes X \boxtimes X \arrow[r, "\mathrm{id} \boxtimes +"] \arrow[d, "\Delta \boxtimes \mathrm{id} \boxtimes \mathrm{id}"] & X \boxtimes X \arrow[dr] & \\
X \boxtimes X \boxtimes X \boxtimes X \arrow[d, "\mathrm{id} \boxtimes \tau \boxtimes \mathrm{id}"] & & X \\
X \boxtimes X \boxtimes X \boxtimes X \arrow[r, "\cdot \boxtimes \cdot"] & X \boxtimes X \arrow[ur, "+"'] &
\end{tikzcd}$$

(and the corresponding one expressing distributivity on the right side) commute.

1.1.3. *a map (in K) $C \boxtimes X \xrightarrow{\cdot} X$ such that*

(i) *$C \boxtimes X \xrightarrow{\cdot} X$ is distributive with respect to $+$ and to the sum of complex numbers (also denoted by $C \boxtimes C \xrightarrow{+} C$). That is, the diagrams*

$$\begin{tikzcd}
C \boxtimes X \boxtimes X \arrow[r, "\mathrm{id} \boxtimes +"] \arrow[d, "\Delta \boxtimes \mathrm{id} \boxtimes \mathrm{id}"] & C \boxtimes X \arrow[dr, "\cdot"] & \\
C \boxtimes C \boxtimes X \boxtimes X \arrow[d, "\mathrm{id} \boxtimes \tau \boxtimes \mathrm{id}"] & & X \\
C \boxtimes X \boxtimes C \boxtimes X \arrow[r, "\cdot \boxtimes \cdot"] & X \boxtimes X \arrow[ur, "+"'] &
\end{tikzcd}$$

and

$$\begin{tikzcd}
C \boxtimes C \boxtimes X \arrow[r, "+ \boxtimes \mathrm{id}"] \arrow[d, "\mathrm{id} \boxtimes \mathrm{id} \boxtimes \Delta"] & C \boxtimes X \arrow[dr, "\cdot"] & \\
C \boxtimes C \boxtimes X \boxtimes X \arrow[d, "\mathrm{id} \boxtimes \tau \boxtimes \mathrm{id}"] & & X \\
C \boxtimes X \boxtimes C \boxtimes X \arrow[r, "\cdot \boxtimes \cdot"] & X \boxtimes X \arrow[ur, "+"'] &
\end{tikzcd}$$

commute.

(ii) *$C \boxtimes X \xrightarrow{\cdot} X$ is associative with respect to $\cdot$ and the product of complex numbers (also denoted $C \boxtimes C \xrightarrow{\cdot} C$). That is, the diagrams*

$$\begin{tikzcd}
C \boxtimes C \boxtimes X \arrow[r, "\mathrm{id} \boxtimes \cdot"] \arrow[d, "\cdot \boxtimes \mathrm{id}"] & C \boxtimes X \arrow[d, "\cdot"] \\
C \boxtimes X \arrow[r, "\cdot"] & X
\end{tikzcd}
\quad \textit{and} \quad
\begin{tikzcd}
C \boxtimes X \boxtimes X \arrow[r, "\mathrm{id} \boxtimes \cdot"] \arrow[d, "\cdot \boxtimes \mathrm{id}"] & C \boxtimes X \arrow[d, "\cdot"] \\
C \boxtimes X \arrow[r, "\cdot"] & X
\end{tikzcd}$$

commute.

(iii) *The action of $1 \xrightarrow{1} C$ is the identity. That is, the diagram*

$$\begin{tikzcd}
X = 1 \boxtimes X \arrow[r, "1 \boxtimes \mathrm{id}"] \arrow[dr, "\mathrm{id}"'] & C \boxtimes X \arrow[d, "\cdot"] \\
& X
\end{tikzcd}$$

commutes.

We will denote such a K-topological algebra by $A = (X, +, ., .)$ and its underlying Kelley space by $X = |A|$.

It is clear that the complex numbers with the ordinary topology and algebraic operations form a K-topological algebra, which, by abuse of language, we will also denote by C.

Part 1.1.1 in the above definition expresses the fact that a K-topological algebra is *an Abelian group* in $\boldsymbol{K}$ (in the sense of [15] exercise 2.c or [13]), or a *weak group* in the sense of [34]. The continuous sum $X \boxtimes X \xrightarrow{+} X$ is in general only continuous on compact subsets when considered as a map $X \times X \xrightarrow{+} X$ and hence the topology of X will not (in general) be a group topology. This notion was introduced by Spanier [34] to obtain some results in algebraic topology (exploiting the clear fact that for arbitrary $Z \in \mathbf{Top}_2$, the identity map $\mathrm{Ke}Z \to Z$ is always a homotopy equivalence). A *different* notion relating group structures and functions continuous on compact subsets has also been considered: Noble in [31, Chap. V] defines a "*k-group*" as being a group X with a group topology behaving within the category of topological groups as compactly generated spaces behave within the category of all Hausdorff

topological spaces. More precisely, a *morphism of groups $X \to G$ into any other topological group is continuous if and only if it is continuous over compact subsets.* The topology of a k-group, however, need not be compactly generated (cf. [31] corollaries to Th. 5.2).

The product (1.1.2 above) of a K-topological algebra will be, in general, continuous only on compact subsets as a map $X \times X \to X$. On the other hand, the product by scalars (1.1.3) is continuous $C \times X \to X$, because C being locally compact, we have $C \boxtimes X = C \times X$ (0.11).

We see then that a K-topological algebra is simply an algebra over the complex numbers with a Hausdorff compactly generated topology which makes the sum and the product continuous when each variable is restricted to a compact subset, and the product by scalars globally continuous.

Any Hausdorff topological algebra with continuous multiplication (in the sense of [17], for instance) determines canonically a K-topological algebra consisting of the same underlying set with the K-ation of the given topology and the same algebraic operations. This is clear since the functor Ke being a right adjoint, preserves limits, and therefore for any Hausdorff spaces Y, Z there is an isomorphism $\mathrm{Ke}(Y \times Z) \approx \mathrm{Ke}Y \boxtimes \mathrm{Ke}Z$. Observe that different topological algebras may determine the same K-topological algebra.

Definition 1.1 is categorical and could have been given in any category with finite products and a terminal object (the empty product). If considered in a category equivalent to $\boldsymbol{K}$, Def. 1.1 would yield the concept of a mathematical object which is categorically indistinguishable from the concept of K topological algebra. Therefore (cf. end of §0) we can think on K-topological algebras as being a Hausdorff space with a structure of complex algebra in which sum, product and product by scalars are continuous only on compact subsets. A morphism is then a linear multiplicative function which is continuous on compact sets. In this approach, any topological algebra in the classical sense with a product continuous on compact sets *is* a K-topological algebra. With this interpretation, however, algebraically isomorphic topological algebras with the same compact subsets are considered equal. Observe that metrizable algebras (in particular, Fréchet algebras, normed algebras) satisfy directly our definition of K topological algebras, since metrizable spaces are in $\boldsymbol{K}$ (see 0.1).

It may be interesting to observe that the K-ation of a topological algebra may fail to be a topological algebra. We have the following:

1.2. Example. Let A be an arbitrary complex vector space of dimension larger than $\aleph_0$, and define a (locally convex, see [22]) topology on A by means of the semi norms $p(a) = |f(a)|$ where f ranges over the set of all linear maps $f\colon A \to C$. If the product on A is defined by $ab = 0$ for all $a, b \in A$, clearly A is a topological (locally m-convex, see [27]) algebra (observe that the continuity of the product is obvious, and therefore we actually don't need to know that A is locally m-convex). It follows from [22] (p. 53, Ex. H), that the compact subsets of A are finite dimensional, whence the topology of $\mathrm{Ke}A$ can be described by: $O \subset \mathrm{Ke}A$ is open if and only if for

every finite dimensional subspace $F \subset A$ it follows that $O \cap F$ is open in F (for the only Hausdorff linear topology of F), i.e., KeA has the *finite topology* of [20]. But then it follows from Th. 1 in [20] that $\mathrm{Ke}A \times \mathrm{Ke}A \xrightarrow{+} \mathrm{Ke}A$ is not continuous, and therefore KeA is *not* a topological algebra (although its product *is* continuous).

Observe that, as in the example above, we can always make a K-topological vector space (namely, an object $X \in K$ together with 1.1.1 and 1.1.3 of Def. 1.1 above) into a K-topological algebra by defining the product ($ab = 0$) as $(X \boxtimes X \to X) = (X \boxtimes X \to 1 \xrightarrow{0} X)$.

The concept of K-topological algebra includes many types of algebras which fail to be topological algebras, and which have been studied under the somewhat artificial concept of topological algebra with partially continuous operations. The following examples also show that K-topological algebras abound in traditional fields such as von Neumann algebras and convolution algebras. The text resumes on page 298.

1.3. Example. Let $\mathcal{H}$ be a complex Hilbert space with inner product $(x \mid y)$ and norm $\|x\| = (x \mid x)^{1/2}$. We denote by $\mathfrak{B}(\mathcal{H})$ the set of all linear bounded operators $T: \mathcal{H} \to \mathcal{H}$. If $T \in \mathfrak{B}(\mathcal{H})$ we denote $\|T\| = \mathrm{Sup}\{\|Tx\|; x \in \mathcal{H}, \|x\| \leqslant 1\}$, and $T^* \in \mathfrak{B}(\mathcal{H})$ will be the adjoint of T, characterized by the identity $(Tx \mid y) = (x \mid T^*y)$ for all $x, y \in \mathcal{H}$. We define $\mathfrak{B}(\mathcal{H}) \times \mathfrak{B}(\mathcal{H}) \xrightarrow{+} \mathfrak{B}(\mathcal{H})$, $\mathfrak{B}(\mathcal{H}) \times \mathfrak{B}(\mathcal{H}) \to \mathfrak{B}(\mathcal{H})$ and $C \times \mathfrak{B}(\mathcal{H}) \xrightarrow{\cdot} \mathfrak{B}(\mathcal{H})$ by $(S+T)x = Sx + Tx$, $(ST)x = S(Tx)$ and $(\lambda T)x = \lambda Tx$. It is clear that $\|\alpha S + \beta T\| \leqslant |\alpha|\,\|S\| + |\beta|\,\|T\|$, $\|ST\| \leqslant \|S\|\,\|T\|$, $\|T^*\| = \|T\|$; $(\alpha S + \beta T)^* = \bar{\alpha}S^* + \bar{\beta}T^*$ and $(ST)^* = T^*S^*$ for all $S, T \in \mathfrak{B}(\mathcal{H})$ and α, β complex ($\bar{\alpha}, \bar{\beta}$ are the conjugates of α, β). In particular $\mathfrak{B}(\mathcal{H})$ is an *algebra over the complex numbers with an involution* $T \rightsquigarrow T^*$. If $\dim \mathcal{H} = n < +\infty$ then $\mathfrak{B}(\mathcal{H})$ is isomorphic to the algebra of $n \times n$ complex matrices. For general facts concerning Hilbert spaces, we refer to [8] or [11].

We will consider now several topologies on $\mathfrak{B}(\mathcal{H})$ that have been extensively used in von Neumann algebras (cf. [8]).

The *uniform topology* on $\mathfrak{B}(\mathcal{H})$ is the topology induced by the norm $\|T\| = \mathrm{Sup}\{\|Tx\|; x \in \mathcal{H}, \|x\| \leqslant 1\}$. With this norm, $\mathfrak{B}(\mathcal{H})$ is a C^*-algebra with identity $I = \mathrm{id}_{\mathcal{H}}$ (cf. [9]).

The *strong topology* on $\mathfrak{B}(\mathcal{H})$ is determined by the following notion of convergence: if $\{T_\alpha\}_{\alpha \in A}$ is a net in $\mathfrak{B}(\mathcal{H})$ and $T \in \mathfrak{B}(\mathcal{H})$, then $T_\alpha \to T$ in the strong topology (or *strongly*) if for each $x \in \mathcal{H}$, we have $\|T_\alpha x - Tx\| \to 0$ with $\alpha \in A$. If $S_\alpha \to S$, $T_\beta \to T$ strongly, it is clear that $\|(S_\alpha + T_\beta)x - (S+T)x\| \leqslant \|(S_\alpha - S)x\| + \|(T_\beta - T)x\| \to 0$ so that $\mathfrak{B}(\mathcal{H}) \times \mathfrak{B}(\mathcal{H}) \xrightarrow{+} \mathfrak{B}(\mathcal{H})$ is continuous for this topology. Also, if $\lambda_\gamma \to \lambda$ in C (= complex numbers), the $\|\lambda_\gamma T_\beta x - \lambda Tx\| \leqslant \|T(\lambda - \lambda)x\| + |\lambda_\gamma|\,\|T_\beta x - Tx\| \to 0$ and $C \times \mathfrak{B}(\mathcal{H})$ is also strongly continuous.

The *ultrastrong topology* on $\mathfrak{B}(\mathcal{H})$ is determined by the following notion of convergence. Let $X = \{x_k\}_{k=1}^{\infty}$ be a sequence in $\mathcal{H}$ with the property

$$(*) \qquad \sum_{k=1}^{\infty} \|x_k\|^2 < +\infty$$

and for $T \in \mathfrak{B}(\mathcal{H})$ define $p_X(T) = [\sum_{k=1}^{\infty} \|Tx_k\|^2]^{1/2}$ (observe that $\sum_{k=1}^{\infty} \|Tx_k\|^2 \leqslant \|T\|^2 \sum_{k=1}^{\infty} \|x_k\|^2$ is finite). A net $\{T_\alpha\}_{\alpha \in A}$ in $\mathfrak{B}(\mathcal{H})$ is convergent to $T \in \mathfrak{B}(\mathcal{H})$ in the ultrastrong topology (or *ultrastrongly convergent*) if $p_X(T_\alpha - T) \to 0$ for each $X = \{x_k\}_{k=1}^{\infty}$ satisfying (*). Since it can be easily verified that $p_X(S+T) \leqslant p_X(S) + p_X(T)$ and $p_X(\lambda T) = |\lambda|\, p_X(T)$, it follows that $\mathfrak{B}(\mathcal{H}) \times \mathfrak{B}(\mathcal{H}) \xrightarrow{+} \mathfrak{B}(\mathcal{H})$ and $C \times \mathfrak{B}(\mathcal{H}) \to \mathfrak{B}(\mathcal{H})$ are ultrastrongly continuous.

The *weak topology* on $\mathfrak{B}(\mathcal{H})$ is determined by the following notions of convergence: a set $\{T_\alpha\}_{\alpha \in A}$ converges in the weak topology (or *weakly*) to T if $(T_\alpha x | y) \to (Tx | y)$ for all $x, y \in \mathcal{H}$. Again if $S_\alpha \to S$, $T_\beta \to T$ weakly, then $((S_\alpha + T_\beta)x | y) = (S_\alpha x | y) + (T_\beta x | y)$ converges to $(Sx | y) + (Tx | y) = ((S+T)x | y)$ and if $\lambda_\gamma \to \lambda$ in C, then $|(\lambda_\gamma T_\alpha x | y) - (\lambda Tx | y)| \to 0$; thus $\mathfrak{B}(\mathcal{H}) \times \mathfrak{B}(\mathcal{H})$ and $C \times \mathfrak{B}(\mathcal{H}) \to \mathfrak{B}(\mathcal{H})$ are weakly continuous.

The *ultraweak topology* on $\mathfrak{B}(\mathcal{H})$ is determined by the following notion of convergence. If $X = \{x_k\}_{k=1}^{\infty}$ and $Y = \{y_k\}_{k=1}^{\infty}$ satisfy (*) above and $T \in \mathfrak{B}(\mathcal{H})$, set $p_{X,Y}(T) = \sum_{k=1}^{\infty} |(Tx_k | y_k)|$ (observe that $\sum_{k=1}^{\infty} |(Tx_k | y_k)| \leqslant \|T\| \sum_{k=1}^{\infty} \|x_k\| \|y_k\| \leqslant \|T\| (\sum_{k=1}^{\infty} \|x_k\|^2)^{1/2} (\sum_{k=1}^{\infty} \|y_k\|^2)^{1/2} < +\infty$). Then, a set $\{T_\alpha\}_{\alpha \in A}$ converges in the ultraweak topology (or ultraweakly) to T if $p_{X,Y}(T_\alpha - T) \to 0$ for all sequences X, Y satisfying (*). We have again $p_{X,Y}(S+T) \leqslant p_{X,Y}(S) + p_{X,Y}(T)$ and $p_{X,Y}(\lambda T) = |\lambda|\, p_{X,Y}(T)$ and therefore $\mathfrak{B}(\mathcal{H}) \times \mathfrak{B}(\mathcal{H}) \xrightarrow{+} \mathfrak{B}(\mathcal{H})$ and $C \times \mathfrak{B}(\mathcal{H}) \to \mathfrak{B}(\mathcal{H})$ are ultraweakly continuous.

It is easy to see that the uniform topology is the strongest and the weak topology the weakest, and in fact we have:

$$\text{uniform} \to \text{ultrastrong} \begin{array}{c} \nearrow \ \text{strong} \ \searrow \\ \searrow \ \text{ultraweak} \ \nearrow \end{array} \text{weak} .$$

If $\dim H = \infty$, all these topologies are different [8]; if $\dim H < \infty$, they all coincide.

A *-subalgebra A of $\mathfrak{B}(\mathcal{H})$ (that is, a subset $\mathsf{A} \subset \mathfrak{B}(\mathcal{H})$ such that whenever $S, T \in \mathsf{A}$, and $\lambda \in C$ we have $S \pm T \in \mathsf{A}$, $\lambda T \in \mathsf{A}$, $T^* \in \mathsf{A}$ and $I \in \mathsf{A}$, where $I = \mathrm{id}_{\mathcal{H}}$) is by definition a *von Neumann algebra* provided it is closed in the weak topology. As a matter of fact, it can be proved (see [8] I.3.4, Th. 2) that a *-subalgebra of $\mathfrak{B}(\mathcal{H})$ is weakly closed if and only if it is closed in either the ultraweak, strong or ultrastrong topologies (but not the uniform topology). Given a von Neumann algebra A, we will denote by A_u, A_s, A_{us}, A_w and A_{uw} the algebra A together with the topologies uniform, strong, ultrastrong, weak and ultraweak, respectively.

Assume now that $K \subset \mathsf{A}$ is compact in any of the above topologies. Then, necessarily, K is weakly compact. This implies that for $x, y \in \mathcal{H}$, the function $T \rightsquigarrow (Tx | y)$ is (continuous, hence) bounded on K. It follows from the principle of uniform boundedness [9] (or using the elementary proof in [18]) that $\operatorname{Sup}\{\|T\|: T \in K\} < +\infty$, that is to say, K is *norm bounded* in $\mathfrak{B}(\mathcal{H})$. With this remark in mind, we can now prove:

1.3.1. $\mathsf{A} \times \mathsf{A} \xrightarrow{+} \mathsf{A}$ *is continuous on compact sets for each of the topologies: uniform, strong and ultrastrong.*

In other words, $S, T \rightsquigarrow ST$ is continuous as a map $\mathrm{Ke}(A_u \times A_u) \to A_u$, $\mathrm{Ke}(A_s \times A_s) \to A_s$ and $\mathrm{Ke}(A_{us} \times A_{us}) \to A_{us}$ and therefore: $\mathrm{Ke}A_u = A_u$, $\mathrm{Ke}A_s$ *and* $\mathrm{Ke}A_{us}$ *are K-topological algebras.*

Consider the case of A_u. The product is continuous everywhere due to the inequality $\|ST\| \leqslant \|S\|\,\|T\|$; moreover, this topology is metrizable (being determined by a single norm) and therefore (from 0.1), $\mathrm{Ke}A_u = A_u$. Observe now that (cf. [8]):

1.3.2. *If $r > 0$ and $B_r \subset \mathfrak{B}(\mathcal{H})$ is the ball $B_r = \{T \in \mathfrak{B}(\mathcal{H});\, \|T\| \leqslant r\}$ then the strong and ultrastrong topologies agree on B_r.* This can be seen as follows. Let $\{T_\alpha\}_{\alpha \in A}$ be a net in B_r converging strongly to $T \in B_r$. For $X = \{x_k\}_{k=1}^\infty$ satisfying (*) above we have $\Sigma_{k=1}^\infty \|(T_\alpha - T)x_k\|^2 = \Sigma_{k=1}^N \|(T_\alpha - T)x_k\|^2 + \Sigma_{k=N+1}^\infty \|(T_\alpha - T)x_k\|^2 \leqslant \Sigma_{k=1}^N \|(T_\alpha - T)x_k\|^2 + 2r\, \Sigma_{k=N+1}^\infty \|x_k\|^2$, so that $\limsup_{\alpha\in A} \Sigma_{k=1}^\infty \|(T_\alpha - T)x_k\|^2 \leqslant \limsup_{\alpha \in A} \Sigma_{k=1}^N \|(T_\alpha - T)x_k\|^2 + 2r\, \Sigma_{k=N+1}^\infty \|x_k\|^2$. But $\lim_{\alpha\in A} \Sigma_{k=1}^N \|(T_\alpha - T)x_k\|^2 = 0$ hence $\limsup_{\alpha \in A} \Sigma_{k=1}^\infty \|(T_\alpha - T)x_k\|^2 \leqslant 2r\, \Sigma_{k=N+1}^\infty \|x_k\|^2$ for all $N = 1, 2, \ldots$, and this means that $p_X(T_\alpha - T) \to 0$; thus the strong and ultrastrong topologies agree on B_r.

The first consequence of this fact is that it is enough to prove the continuity of the product on strongly compact sets. This is done as follows: let $\{S_\alpha\}_{\alpha \in A}$ and $\{T_\beta\}_{\beta \in B}$ be nets such that $S_\alpha \to S$, $T_\beta \to T$ strongly and $\|S_\alpha\| \leqslant r$, $\|T_\beta\| \leqslant r$ for all α, β and some $r > 0$. Assume $X = \{x_k\}_{k=1}^\infty$ satisfies (*) above. Then $p_X(S_\alpha T_\beta - ST) \leqslant p_X(S_\alpha(T_\beta - T)) + p_X((S_\alpha - S)T)$. Now, for all $U, V \in \mathfrak{B}(\mathcal{H})$ it is easy to see that $p_X(UV) \leqslant \|U\|\, p_X(V)$ and $p_X(UV) \leqslant \|V\|\, p_X(U)$. Then $p_X(S_\alpha T_\beta - ST) \leqslant \|S_\alpha\|\, p_X(T_\beta - T) + \|T\|\, p_X(S_\alpha - S) \leqslant r p_X(T_\beta - T) + \|T\|\, p_X(S_\alpha - S) \to 0$ with $\alpha \in A$, $\beta \in B$. Therefore, $S, T \rightsquigarrow ST$ is strongly (eq., ultrastrongly) continuous as a map $B_r \times B_r \to B_{r^2}$, and our claim follows.

The second consequence of the agreement on normed bounded sets of the strong and ultrastrong topologies is that they have the same K-ation: for any von Neumann algebra A, $\mathrm{Ke}A_s = \mathrm{Ke}A_{us}$; this means that A_s and A_{us} are the same K-topological algebra.

It should be remarked that $\mathrm{Ke}A_s = \mathrm{Ke}A_{us}$ is a genuine K-topological algebra in the sense that it is not obtained as the K-ation of a topological algebra with continuous product. In fact:

1.3.3. *If $A = \mathfrak{B}(\mathcal{H})$ with $\dim \mathcal{H} = \infty$, the product is not continuous as a map $A_w \times A_w \to A_w$, $A_s \times A_s \to A_s$, $A_{uw} \times A_{uw} \to A_{uw}$ or $A_{us} \times A_{us} \to A_{us}$* (cf. [36] or [8] I.3, Ex. 2). This can be seen as follows: Assume $\mathcal{H}$ is separable. If $X = \{x_k\}_{k=1}^\infty$ satisfies (*) above with $c = \Sigma_{k=1}^\infty \|x_k\|^2$, define $B = \{T \in \mathfrak{B}(\mathcal{H});\, p_X(T) \leqslant 1\}$. Let now $\{e_l\}_{l=1}^\infty$ be an orthonormal basis for $\mathcal{H}$. For all N, M positive integers we have $\Sigma_{l=1}^N (\Sigma_{k=1}^M |(x_k | e_l)|^2) \leqslant \Sigma_{k=1}^M (\Sigma_{l=1}^\infty |(x_k | e_l)|^2) = \Sigma_{k=1}^M \|x_k\|^2 \leqslant c$ and therefore $\Sigma_{l=1}^\infty (\Sigma_{k=1}^\infty |(x_k | e_l)|^2) \leqslant c$ which shows that $\lim_{l \to \infty} \Sigma_{k=1}^\infty |(x_k | e_l)|^2 = 0$. Choose $\{d_l\}_{l=1}^\infty$ a sequence of nonnegative reals such that $d_l^2\, \Sigma_{k=1}^\infty |(x_k | e_l)|^2 \leqslant 1$ for $l = 1, 2, \ldots$

and $\lim_{l\to\infty} d_l = \infty$. Let $T_n \in \mathfrak{B}(\mathcal{H})$, $n = 1, 2, \ldots$ be defined by $T_n e_l = \delta_{nl} d_n e_n$ (δ_{nl} = Kronecker's delta). Clearly $\Sigma_{k=1}^{\infty} \| T_k x_n \|^2 = \Sigma_{k=1}^{\infty} \| \Sigma_{l=1}^{\infty} (x_k \mid e_l)\, T_n e_l \|^2 = \Sigma_{k=1}^{\infty} |(x_k \mid e_n)|^2 d_n^2 \leq 1$, i.e., $T_n \in B$ and $\| T_n \| = d_n \to \infty$. This means that $\operatorname{Sup} \{ \| T \|;\ T \in B \} = \infty$, and therefore by the uniform boundedness principle (see [11]) there is a $z \in \mathcal{H}$ such that $\operatorname{Sup} \{ \| Tz \|;\ T \in B \} = \infty$, with (necessarily) $z \neq 0$. Now we will prove that, *for $x, y \in \mathcal{H}$ both different from zero, no choice of $X = \{ x_k \}_{k=1}^{\infty}$ will make true that if $S, T \in B$ then* $|(STx \mid y)| \leq 1$. In fact, if z satisfies

$$(**) \qquad \operatorname{Sup} \{ \| Tz \|;\ T \in B \} = \infty ,$$

define $S \in \mathfrak{B}(\mathcal{H})$ by $Su = \lambda (u \mid x) z$ where $\lambda \in C$ and $|\lambda| > 0$ is small enough in order that $p_X(S) \leq 1$ or $S \in B$. Then choose $T \in B$ with $|\lambda|\, \| y \|\, \| TSx \| = \| y \|\, |\lambda|^2\, \| x \|^2 \| Tz \| > 1$ (use (**) above) and $U \in \mathfrak{B}(\mathcal{H})$ unitary and such that $U^* y = \mu TSx$ where $\mu \in C$ satisfies $|\mu|\, \| TSx \| = \| y \|$. Then $UT, S \in B$ and $|(UTSx \mid y)| = |(TSx \mid U^* y)| = |\lambda \mu|\, \| TSx \|^2 > 1$, as claimed. This shows that $S, T \rightsquigarrow ST$ is discontinuous at $S = T = 0 \in \mathfrak{B}(\mathcal{H})$ as a map $\mathfrak{B}(\mathcal{H})_{us} \times \mathfrak{B}(\mathcal{H})_{us} \to \mathfrak{B}(\mathcal{H})_w$ which implies that the product is discontinuous in all four the weak, strong, ultraweak, or ultrastrong topologies, as claimed in 1.3.3.

An argument similar to the one given above (1.3.2) shows that $\mathrm{Ke} A_w = \mathrm{Ke} A_{uw}$. But in general, the product is *not* continuous on weakly compact sets, in fact not even sequentially continuous, as the following example shows: take $\mathcal{H} = l^2(\boldsymbol{Z})$ where $\boldsymbol{Z} = \{0, \pm 1, \pm 2, \ldots\}$ is the integer and let $U \in \mathfrak{B}(\mathcal{H})$ be the shift defined by $U\{c_k\} = \{d_k\}$ where $d_k = c_{k+1}$, $k = 0, \pm 1, \pm 2, \ldots$. Then $U^n \to 0$, $U^{-n} \to 0$ weakly as $n \to \infty$. However, $U^n U^{-n}$ = Identity.

We see then that some weak von Neumann algebras are K-topological algebras (the finite dimensional algebras, for instance) while some others ($\mathfrak{B}(\mathcal{H})_w$, $\dim H = \infty$) are not. The following important algebra is another example of the former, as we shall see. Let $\boldsymbol{S}$ be a (fixed) set, $\mathcal{H} = l^2(\boldsymbol{S})$ the Hilbert space of all complex functions $x\colon \boldsymbol{S} \to C$ such that $\Sigma \{ |x(s)|^2, s \in \boldsymbol{S} \} < +\infty$ (with operations defined pointwise and inner product $(x \mid y) = \Sigma \{ x(s) \overline{y(s)};\ s \in \boldsymbol{S} \}$). Let $l^\infty(\boldsymbol{S}) \subset \mathfrak{B}(\mathcal{H})$ be the algebra consisting of all the operators on $l^2(\boldsymbol{S})$ expressible as $Tx = dx$ for some $d\colon \boldsymbol{S} \to C$ satisfying $\operatorname{Sup} \{ |d(s)|;\ s \in \boldsymbol{S} \} < +\infty$ and where dx is the function $(dx)(s) = d(s) x(s)$. It is not hard to see (cf. [8] I.7) that $l^\infty(\boldsymbol{S})$ is a von Neumann algebra which is ***Abelian*** (i.e. $ST = TS$ for $S, T \in l^\infty(\boldsymbol{S})$). We shall identify $T \in l^\infty(\boldsymbol{S})$ with $d\colon \boldsymbol{S} \to C$ when $Tx = dx$ for all $x \in l^2(\boldsymbol{S})$. It is clear that $\| T \| = \operatorname{Sup} \{ |d(s)|;\ s \in \boldsymbol{S} \}$, $T^* = \bar{d}$ the conjugate function, and $(dd')(s) = d(s)\, d'(s)$. It follows from ([8] III.6, Prop. 5, 6 and 7) that $l^\infty(\boldsymbol{S})_w = l^\infty(\boldsymbol{S})_{uw}$ and $l^\infty(\boldsymbol{S})_s = l^\infty(\boldsymbol{S})_{us}$ (in fact, with the notation of [8], $C_A = 1$ when $A = l^\infty(\boldsymbol{S})$). In other words, the weak (resp., strong) topology coincides on $l^\infty(\boldsymbol{S})$ with the ultraweak (resp., ultrastrong). Actually one can prove in a very elementary way that the weak and ultraweak topologies on $l^\infty(\boldsymbol{S})$ both agree with the weak * topology of $l^\infty(\boldsymbol{S})$ as the Banach dual of $l^1(\boldsymbol{S})$ (cf. [11]), which means that on $l^\infty(\boldsymbol{S})$ the weak or ultraweak convergence is determined by the seminorms $x \rightsquigarrow |\Sigma \{ x(s) z(s);\ s \in \boldsymbol{S} \}|$ where $z \in l^1(\boldsymbol{S})$, i.e., $\Sigma \{ |z(s)|;\ s \in \boldsymbol{S} \} < +\infty$. With this in mind, one can prove independently of w = us, s = us, that

1.3.4. *If $K \subset l^\infty(S)$, the following are equivalent:*

(i) *K is strongly (= ultrastrongly) compact,*

(ii) *K is ultraweakly compact,*

(iii) *K is strongly closed and norm bounded,*

(iv) *K is ultraweakly closed and norm bounded,*

(v) *K is weakly closed and norm bounded.*

Moreover, if K satisfies any of the above, the weak, ultraweak, strong (and ultrastrong) topologies coincide on K.

We will indicate a proof of 1.3.4 based on the plan: (iii) $\Longleftrightarrow$ (i) $\Rightarrow$ (ii) $\Longleftrightarrow$ (iv) and (ii) $\Rightarrow$ (v) $\Rightarrow$ (i). The following remarks are elementary: first, the equivalence (ii) $\Longleftrightarrow$ (iv) is the Alaoglu-Bourbaki Theorem [11]; second, we observed above (last paragraph before 1.3.1) that weakly (hence ultraweakly, strongly or ultrastrongly) compact implies norm bounded, whence (i) $\Rightarrow$ (iii). On normed bounded sets, the strong and ultrastrong topologies agree (1.3.2 above) and therefore the strong topology is stronger on these than the ultraweak, which proves (i) $\Rightarrow$ (ii). For similar reasons, (ii) $\Rightarrow$ (v). The implications (iii) $\Rightarrow$ (i) and (v) $\Rightarrow$ (i), together with the second half of 1.3.4 follow from the following: if $\{d_\alpha\}_{\alpha\in A}$ is a net, $\|d_\alpha\| \leqslant r$ and $d_\alpha(s) \to 0$ for each $s \in S$, then $d_\alpha \to 0$ strongly. In fact, for each $x \in l^2(S)$ and a finite $F \subset S$, we have $\|d_\alpha x\|^2 \leqslant \Sigma_{s\in F} |d_\alpha(s)\,x(s)|^2 + r\Sigma_{s\notin F} |x(s)|^2$, whence $\limsup_{\alpha\in A} \|d_\alpha x\|^2 \leqslant \Sigma_{s\notin F} |x(s)|^2$ and since F is arbitrary, $\lim_{\alpha\in A} \|d_\alpha x\|^2 = 0$ as claimed.

Another way of writing 1.3.4 is the following:

1.3.5. *For any set S,*

$$\mathrm{K}e\,l^\infty(S)_{\mathrm{w}} = \mathrm{K}e\,l^\infty(S)_{\mathrm{uw}} = \mathrm{K}e\,l^\infty(S)_{\mathrm{s}} = \mathrm{K}e\,l^\infty(S)_{\mathrm{us}}\,.$$

It is interesting to observe that *the product $S, T \rightsquigarrow ST$ is not weakly (= ultraweakly) or strongly (= ultrastrongly) continuous, even when restricted to $l^\infty(S)$,* when S is infinite. In fact, we can assume that $S \supset \{1, 2, \ldots\} = N$. Let $h \in l^2(S)$ be defined by $h(n) = 1/n$, $n \in N$, $h = 0$ elsewhere. Consider $V = \{d \in l^\infty(S);\ \|dh\| \leqslant 1\}$. V is a strong neighborhood of $0 \in l^\infty(S)$. We will see that for *no* strong neighborhood W of $0 \in l^\infty(S)$ it will be true that $d^2 \in V$ if $d \in W$. Clearly, W can be assumed to be of the form $W = \{d;\ \Sigma_{k=1}^\infty \|dx_k\|^2 < \epsilon\}$ for some sequence $\{x_k\}_{k=1}^\infty$ in l^2 satisfying ((*) above): $\Sigma \|x_k\|^2 < +\infty$. Since $\Sigma_{k,n=1}^\infty |x_k(n)|^2 < \infty$ we conclude that $\{\Sigma_{k=1}^\infty |x_k(n)|^2\}_{n=1}^\infty$ is summable (this is the same argument used in the proof of 1.3.3). Therefore for some positive integer m we have $\epsilon^{-1}\Sigma_{k=1}^\infty |x_k(n)|^2 < m^{-1}$. Choose λ real such that $\epsilon^{-1}\Sigma_{k=1}^\infty |x_k(n)|^2 < \lambda^{-2} < m^{-1}$ and define $d_0 \in l^\infty(S)$ by $d_0(m) = \lambda$, $d_0 = 0$ elsewhere. Then $\Sigma_{k=1}^\infty \|d_0 x_k\|^2 = \Sigma_{k=1}^\infty \lambda^2 |x_k(n)|^2 = \lambda^2 \Sigma_{k=1}^\infty |x_k(n)|^2 < \epsilon$ and $\|d_0^2 h\| = \lambda^2 m^{-1} > 1$ so that $d_0 \in W$ and $d_0^2 \notin V$, as claimed. This shows that $d \rightsquigarrow d^2$ is not continuous as a map $l^\infty(S)_{\mathrm{us}} \to l^\infty(S)_{\mathrm{s}}$. A similar argument shows that $d \rightsquigarrow d^2$ is not continuous as a map $l^\infty(S)_{\mathrm{uw}} \to l^\infty(S)_{\mathrm{w}}$, and therefore the product in $l^\infty(S)$ is not continuous for any of the topologies weak, ultraweak, strong, ultrastrong.

Actually, this proof yields a bit more: if $P = \{d \in l^\infty(S), d(s) \geqslant 0\}$ and P_w $(= P_{uw})$, P_s $(= P_{us})$ denote P with the relative weak (= ultraweak) and strong (= ultrastrong) topologies, respectively and $\varphi: P \to P$ is the map $\varphi(d) = d^2$, then we just saw the proof of the first half of:

1.3.6. *If S is infinite:*
 (i) *φ is not continuous as a map $P_w \to P_w$, $P_s \to P_s$ $(P_{uw} \to P_{uw}, P_{us} \to P_{us})$*
 (ii) *$P_s \xrightarrow{\varphi} P_w = P_{us} \xrightarrow{\varphi} P_{uw}$ is a homeomorphism.*

It is clear that 1.3.6 (ii) follows from $\Sigma d^2(s)|x(s)y(s)| = \Sigma(d(s)|x(s)y(s)|^{1/2})^2$ for $x, y \in l^2(S)$.

Observe that if we assume that $\mathrm{Ke}P_w = P_w$, then $P_s \xrightarrow{\varphi} P_w = \mathrm{Ke}P_w = \mathrm{Ke}P_s \xrightarrow{\mathrm{id}} P_s$ would be continuous contradicting 1.3.6 (i) when S is infinite. Thus:

1.3.7. *If S is infinite, then* $\mathrm{Ke}\, l^\infty(S)_w \neq l^\infty(S)_w$.

In fact, it can be proved that when S is countable, $\mathrm{Ke}\, l^\infty(S)_w$ coincides with the relative product topology of $l^\infty(S) \subset \prod_S C$. This completes our Example 1.3.

1.4. Example. For general results concerning topological groups we refer the reader to [2] and [3]. Let G be a Hausdorff locally compact group and denote by $U(G)$ the algebra of all bounded uniformly continuous functions on G with values in the complex numbers C. The norm $\|f\|_\infty = \mathrm{Sup}\{|f(s)|; s \in G\}$ makes $U(G)$ a Banach algebra (in fact a C^*-algebra since it is easy to see that $U(G)$ is a closed $*$-subalgebra of the C^*-algebra $B(G)$ of all bounded complex valued function on G with the norm $\|f\|_\infty$). For $s \in G$ and $f \in U(G)$, the function $\gamma(s)f$ is defined by $[\gamma(s)f](t) = f(s^{-1}t)$. Clearly $\gamma(s)f \in U(G)$ and $\|\gamma(s)f\|_\infty = \|f\|_\infty$. Thus G acts (isometrically) on $U(G)$ by $s \rightsquigarrow \gamma(s)$. If $S \subset U(G)$ is any subset, we denote by $\Lambda(S)$ the linear subspace of $U(G)$ spanned by S. If $f \in U(G)$, then $[f] \subset U(G)$ denotes the subspace $[f] = \Lambda\{\gamma(f)f; s \in G\}$ generated by the (left) translates of f. We shall abbreviate $\dim[f] = \dim_C[f]$.

1.4.1. $\dim[f]$ *is finite for all $f \in U(G)$ if and only if G is finite.*

Proof. It is clear that if G is finite, $\dim_C U(G) = \mathrm{card}(G) < +\infty$ and therefore $\dim[f] \leqslant \mathrm{card}(G)$ for all $f \in U(G)$.

Assume now that $\dim[f] < +\infty$ for all $f \in U(G)$ and define $U_n = \{f \in U(G): \dim[f] \leqslant n\}$ for $n = 1, 2, \ldots$. Clearly $U(G) = \bigcup_{n=1}^\infty U_n$. We claim that each U_n is closed in $U(G)$. In fact, assume $f_j \to f$ (as $j \to +\infty$) in $U(G)$ and $f_j \in U_n$ for $j = 1, 2, \ldots$. Then for any choice of $n+1$ elements $s_0, s_1, \ldots, s_n$ of G the functions $\gamma(s_0)f_j, \ldots, \gamma(s_n)f_j$ are linearly dependent and therefore there are complex numbers $\alpha_0^j, \alpha_1^j, \ldots, \alpha_n^j$ such that $\Sigma_{k=0}^n \alpha_k^j \gamma(s_k)f_j = 0$ and $\Sigma_{k=0}^n |\alpha_k^j| = 1$ for each $j = 1, 2, \ldots$. By passing to an appropriate subsequence we can assume that $\alpha_k^j \to \alpha_k$ as $j \to \infty$ for each $k = 0, 1, \ldots, n$ and therefore $\Sigma|\alpha_k| = 1$ also. But clearly from $f_j \to f$ and $\alpha_k^j \to \alpha_k$

(as $j \to \infty$) we obtain $\alpha_k^j \gamma(s_k) f_j \to \alpha_k \gamma(s_k) f$ in $U(G)$ and therefore $\sum_{k=0}^n \alpha_k \gamma(s_k) f = \lim_{j\to\infty} \sum_{k=0}^n \alpha_k^j \gamma(s_k) f_j = 0$ which shows that any $n+1$ translates of f are linearly dependent or in other words, that $f \in U_n$. (Observe that we have actually proved that U_n is closed in $U(G)$ for *any* linear, translation invariant topology on $U(G)$.) Since $U(G)$ under $\|f\|_\infty$ is a complete metric space, from Baire's Theorem (see [21]) follows that there exist $n_0, f_0 \in U_{n_0}$ and $\epsilon > 0$ such that if $h \in U(G)$ and $\|h\|_\infty \leq \epsilon$, then $h - f_0 \in U_{n_0}$, i.e., $\dim[h - f_0] \leq n_0$. Observe now that for $f, f' \in U(G)$, λ a nonzero scalar, always $\dim[f + f'] \leq \dim[f] + \dim[f']$ and $\dim[\lambda f] = \dim[f]$. Hence, for $f \in U(G)$ and λ small enough (so that $\lambda \|f\|_\infty < \epsilon$) we have $\dim[f] = \dim[\lambda f] \leq \dim[\lambda f - f_0] + \dim[f_0] \leq 2n_0$. Thus $U(F) = U_{2n_0}$. But now if $s_1, \ldots, s_m$ are distinct elements of G, there is a compact neighborhood of V of the identity of G such that the sets $s_j V$, $j = 1, \ldots, m$ are pairwise disjoint. Let $f \in U(G)$ be a function with support in $s_1 V$ and satisfying $f(s_1) = 1$. Clearly $\dim[f] \geq m$. It follows that $m \leq 2n_0$ and therefore G cannot have more than $2n_0$ different elements, or, G is finite.

We recall that if f is a function on a locally compact group G and μ is a measure on G, the *convolution* $\mu * f$ is the function

$$(\mu * f)(x) = \int f(s^{-1}x)\, d\mu(s)$$

and in particular, if ϵ_t is the point mass measure at $t \in G$ with total mass $+1$, then

$$(\dagger) \qquad \epsilon_t * f = \gamma(t) f;$$

if μ and ν are measures on G, $\mu * \nu$ is the measure satisfying

$$\int f(x)\, d(\mu * \nu)(x) = \iint_{G \times G} f(sx)\, d\mu(s)\, d\nu(x)$$

for each f, say, continuous with compact support. Fubini's Theorem applies to show that

$$(\dagger\dagger) \qquad (\mu * \nu)(f) = \nu(\check{\mu} * f)$$

where for any function h and measure β, we write $\beta(h) = \int h(x)\, d\beta(x)$ and $\check{\beta}(h) = \int h(x^{-1})\, d\beta(x)$. A table of sufficient conditions for the existence of $\mu * f$ and $\mu * \nu$ can be found in the last page of [3]. One of these is the following: *if μ and ν are bounded then $\mu * \nu$ exists and is bounded* (we recall that a measure β is *bounded* if

$$\|\beta\| = \operatorname{Sup}\{|\beta(f)| : f \in K(G),\ |f(s)| \leq 1 \text{ for all } s \in G\}$$

is finite: $\|\beta\| < +\infty$, where we denote by $K(G)$ the space of continuous functions with compact support). It is not hard to prove that $\mu * \nu$ has desirable properties and in particular that the set of bounded measures under ordinary sum and convolution is an algebra which we will denote by $M^1(G)$. The *vague topology* denoted T_1) on $M^1(G)$ is the topology corresponding to the simple convergence on $K(G)$, i.e., $\mu_\alpha \to \mu$ *vaguely* if $\mu_\alpha(f) \to \mu(f)$ for each $f \in K(G)$. If G is compact then all measures

are bounded and $M^1(G)$ can be identified with the dual of the Banach space $K(G)$ under the norm $\|f\| = \mathrm{Sup}\{|f(s)| \, s \in G\}$, the vague topology coinciding with the w^*-topology. In particular, the vaguely relatively compact sets coincide with the norm-bounded sets of $M^1(G)$. Denote by M_I the algebra $M^1(G)$ endowed with the vague topology T_I.

1.4.2. *Let G be a compact group. The convolution $M_\mathrm{I} \times M_\mathrm{I} \xrightarrow{*} M_\mathrm{I}$ is continuous if and only if G is finite.*

Proof. If G is finite, $M^1(G)$ is finite dimensional and therefore any bilinear map $M_\mathrm{I} \times M_\mathrm{I} \to M_\mathrm{I}$ is continuous. Assume now that G is infinite. First, let us observe that a neighborhood base of $0 \in M_\mathrm{I}$ for the vague topology is provided by the sets $V = \{\mu \in M_\mathrm{I}; |\mu(f_j)| \leqslant \epsilon, j = 1, 2, \ldots, n\}$ where $\epsilon > 0$ and $f_1, f_2, \ldots, f_n \in K(G)$. Let now $f \in K(G)$ be such that $\dim[f] = \infty$ (cf. 1.4.1 above). We are going to show that for any choice of functions $f_1, \ldots, f_n, g_1, \ldots, g_m$ in $K(G)$, there are measures μ, ν such that $\mu(f_j) = 0$, $\nu(g_k) = 0$ for all $1 \leqslant j \leqslant n$, $1 \leqslant k \leqslant m$ and yet $(\mu * \nu)(f) \neq 0$. This will show that $\mu, \nu \rightsquigarrow \mu * \nu$ is not continuous at 0. According to (+) above, $\{\mu * f, \mu \in M_\mathrm{I}\} \supset [\gamma(s) f; s \in G]$ and in particular the linear map $\mu \rightsquigarrow \mu * f$, $M_\mathrm{I} \to K(G)$ has infinite dimensional range. Clearly the subspace $N \subset M_\mathrm{I}$ of all measures satisfying $\mu(f_j) = 0$, $j = 1, 2, \ldots, n$ has finite codimension in M_I, and therefore the linear map $\mu \rightsquigarrow \mu * f$ restricted to $N \to K(G)$ also has infinite dimensional range, which we denote by $R \subset K(G)$. Hence the subspace $[g_1, \ldots, g_m]$ generated by $g_1, \ldots, g_m$ can not contain R. It follows then from the Hahn-Banach Theorem [11] that there is an element $\nu \in (K(G))' = M_\mathrm{I}$ vanishing on $[g_1, \ldots, g_m]$ and such that $\nu(h) \neq 0$ for some $h \in R$. But then necessarily $h = \check{\mu} * f$ for some μ and (cf. (++) above) $(\mu * \nu)(f) = \nu(\check{\mu} * f) \neq 0$, as desired.

1.4.3. *Let G be a compact group. Then the convolution* $\mathrm{Ke}M_\mathrm{I} \otimes \mathrm{Ke}M_\mathrm{I} \xrightarrow{*} \mathrm{Ke}M_\mathrm{I}$ *is continuous. In other words,* $(M_\mathrm{I}, +, *)$ ***is a K-topological algebra.***

Proof. In fact, assume $\mu_\alpha \to 0$, $\nu_\beta \to 0$ vaguely and $\|\mu_\alpha\| \leqslant L$, $\|\nu_\beta\| \leqslant L$ for some L and all α, β. Let $f \in K(G)$ and define $f_\alpha = \check{\mu}_\alpha * f$, or $f_\alpha(x) = \int f(sx)\, d\mu_\alpha(s)$. It is clear that $f_\alpha(x) \to 0$ for each $x \in G$. We shall prove that the family $\{f_\alpha\}$ is equi-uniformly continuous on G. First, for each neighborhood V of the identity $e \in G$, define $V^G = \text{closure} \cup \{xVx^{-1}\}$. Clearly $V \subset V^G$ and V^G is compact. Assume now $z \in V^G$ for all V. Then one can pick $x_V \in G$, $y_V \in V$ such that $x_V y_V x_V^{-1} \in zV$ for each V, for that, in particular, $x_V y_V x_V^{-1} \to z$ following the filter $\{V\}$. G being compact, there is a subnet $\{x_U\}$ of $\{x_V\}$ such that $x_U \to x$ for some $x \in G$. Hence $x_U^{-1} \to x^{-1}$ and since $y_V \in V$, clearly $y_U \to e$. Thus $z = \lim_U x_U y_U x_U^{-1} = xex^{-1} = e$. We conclude that $\cap V^G = \{e\}$. It follows easily that the family $\{V^G\}$ is also a neighborhood base of e. We go back now to the equi-uniform continuity of $\{f_\alpha\}$. Assume $\epsilon > 0$ and choose V such that if $xy^{-1} \in V^G$, then $|f(x) - f(y)| \leqslant \epsilon/L$. This is always possible because f is continuous and G is compact. Assume now that $xy^{-1} \in V$. Then

$|f_\alpha(x) - f_\alpha(y)| \leq \int |f(sx) - f(sy)| \,\mathrm{d}\, |\mu_\alpha|(s) \leq L \operatorname{Sup}\{|f(sx) - f(sy)|; s \in G\}$. But clearly $(sx)(sy)^{-1} = sxy^{-1}s^{-1} \in sVs^{-1} \subset V^G$, so that $\operatorname{Sup}\{|f(sx) - f(sy)|; s\in G\} \leq \epsilon/L$ and therefore $|f_\alpha(x) - f_\alpha(y)| \leq \epsilon$ as desired. Finally, it is easy to see that if $f_\alpha(x) \to 0$ for each x and the family is equicontinuous, then $f_\alpha \to 0$ uniformly on G. Hence $|\nu_\beta(f_\alpha)| \leq \|f_\alpha\| \, \|\nu_\beta\| \leq \|f_\alpha\| \, L \to 0$ and $(\mu_\alpha * \nu_\beta)(f) = \nu_\beta(f_\alpha) \to 0$ as desired. Q.E.D.

There are several interesting variations on this theme. For instance, in the case of a locally compact group G, one can consider the following topologies on $M^1(G)$:

T_{II}. the weak topology of $M^1(G)$ as a Banach space under the norm $\|\beta\|$ defined above.

T_{III}. the topology for which $\mu_\alpha \to \mu$ if and only if

$$\int f(x)\,\mathrm{d}\mu_\alpha(x) \to \int f(x)\,\mathrm{d}\mu(x) \qquad \text{for all } f: G \to C$$

continuous and bounded.

T_{IV} the topology for which $\mu_\alpha \to \mu$ if and only if $\mu_\alpha(U) \to \mu(U)$ for each open set $U \subset G$.

The convolution product is not continuous on either of the above topologies. Yet, it is continuous on compact sets in all cases. In other words, if M_{II}, M_{III} and M_{IV} denotes the algebra $M^1(G)$ with the topologies T_{II}, T_{III} and T_{IV} respectively, then $\mathrm{Ke}M_q$ is a K-topological algebra, $\mathrm{I} \leq q \leq \mathrm{IV}$ (cf. [3] Chap. VIII, §3, Ex. 11). In fact $\mathrm{Ke}M_{\mathrm{III}} = \mathrm{Ke}M_{\mathrm{IV}}$ although $M_{\mathrm{III}} \neq M_{\mathrm{IV}}$. This completes our Example 1.4.

By a morphism of K-topological algebras we will understand a continuous function which is linear and multiplicative. Specifically:

1.5. Definition. *Given two K-topological algebras A, B, a morphism $A \xrightarrow{\varphi} B$ is a map $|A| \xrightarrow{\varphi} |B|$ in **K** such that the diagrams*

$$\begin{tikzcd} {|A| \boxtimes |A|} \arrow[r, "\varphi\boxtimes\varphi"] \arrow[d, shift right, "\cdot"'] \arrow[d, shift left, "+"] & {|B| \boxtimes |B|} \arrow[d, shift right, "\cdot"'] \arrow[d, shift left, "+"] \\ {|A|} \arrow[r, "\varphi"] & {|B|} \end{tikzcd} \quad \textit{and} \quad \begin{tikzcd} C \boxtimes |A| \arrow[r, "\mathrm{id}\boxtimes\varphi"] \arrow[d, "\cdot"] & C \boxtimes |B| \arrow[d, "\cdot"] \\ {|A|} \arrow[r, "\varphi"] & {|B|} \end{tikzcd}$$

commute.

The class of K-topological algebras with the above morphisms between them form a category that we will denote $\boldsymbol{A}$. Given $A, B \in \boldsymbol{A}$, $\boldsymbol{A}_0(A,B)$ will denote the set of morphisms from A to B. Clearly $\boldsymbol{A}_0(A,B) \subset \boldsymbol{K}_0(|A|, |B|)$ and we have a functor $\boldsymbol{A} \xrightarrow{|\;|} \boldsymbol{K}$, the *"underlying Kelley space"* functor. If $A \xrightarrow{\varphi} B$ in $\boldsymbol{A}$, then $|\varphi| = \varphi$.

1.6. Proposition. *$\boldsymbol{A}$ is a **K**-category in such a way that $\boldsymbol{A} \xrightarrow{|\;|} \boldsymbol{K}$ is a **K**-functor. Furthermore, $\boldsymbol{A}(A,B) \xrightarrow{|\;|} \boldsymbol{K}(|A|, |B|)$ is a full injection (cf. 0.6).*

Proof. Define $A(A,B)$ to be the K-ation of $A_0(A,B)$ (considered as a subspace of $K(|A|,|B|)$). The proof then is completely straightforward. For example, the composition $A(A,B) \boxtimes A(B,D) \to A(A,D)$ in K is defined in the diagram

$$\begin{tikzcd} A(A,B) \boxtimes A(B,D) \arrow[r, "\gamma"] \arrow[d, dashed, "|\,|\boxtimes|\,|"] & A(A,D) \arrow[d, dashed, "|\,|"] \\ K(|A|,|B|) \boxtimes K(|B|,|D|) \arrow[r, dashed, "k"] & K(|A|,|D|) . \end{tikzcd}$$

Since k is continuous (see 0.13), it follows that γ is continuous (use the fact that $A(A,D) \overset{|\,|}{\to} K(|A|,|D|)$ is a full injection), that is, $\gamma \in K$. The commutativity of the diagram above is precisely one of the conditions of K-functoriality. Etc. ... Q.E.D.

Observe that $A_0(A,B)$ is a closed subset of $K(|A|,|B|)$, and hence, the topology of $A(A,B)$ is actually the relative topology (see 0.4).

An identity for the product in a K-topological algebra is a map $1 \overset{e}{\to} |A|$ *in K such that the diagrams:*

$$\begin{tikzcd} {|A| \boxtimes |A|} \arrow[drr, dashed] & {|A| \boxtimes 1 = |A| = 1 \boxtimes |A|} \arrow[l, "\mathrm{id}\boxtimes e"'] \arrow[r, "e\boxtimes \mathrm{id}"] \arrow[d, "\mathrm{id}"] & {|A| \boxtimes |A|} \arrow[dl, dashed] \\ & {|A|} & \end{tikzcd}$$

commutes. If the product of A has a unit, we will say that A is an *algebra with identity*. Given two algebras with identity, a morphism in A which preserves the identity in the sense that the diagram

$$\begin{tikzcd} {|A|} \arrow[rr] & & {|B|} \\ & 1 \arrow[ul, "e"] \arrow[ur, "e"'] & \end{tikzcd}$$

commutes, will be called a *morphism of algebras with identity*. Algebras with identity and morphisms of algebra with identity form a (not full) subcategory of A that will be denoted by A', and we have $A'_0(A,B) \subset A_0(A,B)$. Proposition 1.6 holds similarly. In general, we have:

1.7. Proposition. *Let C be a subcategory of A (i.e., a class of K-topological algebras with certain morphisms of K-topological algebras between them, containing all the identities and closed under composition). Then C is a K-category and* $C \overset{|\,|}{\to} K$ *is a K-functor such that for all* $A, B \in C$, $C(A,B) \overset{|\,|}{\to} K(|A|,|B|)$ *is a full injection.*

Proof. Similar to the proof of Proposition 1.6. Q.E.D.

We deduce from this that some standard classes, for example, commutative K-topological algebras, normed (or Banach) algebras, K-topological algebras with involution and morphisms preserving the involution, C^*-algebras, locally multiplicative K-topological algebras, Fréchet algebras, etc., etc., are all $\boldsymbol{K}$-categories.

§ 2. Categorical properties of $\mathcal{A}$ and $\mathcal{A}'$

In this section we will show that $\boldsymbol{A}$ and $\boldsymbol{A}'$ are $\boldsymbol{K}$-complete $\boldsymbol{K}$-categories (cf. [10]). This property furnishes the basic (and only!) tool needed for the duality theory developed in §3. A second important property to be established is the existence of the free $\boldsymbol{K}$-topological algebra over a Kelley space. Furthermore, we will show that $\boldsymbol{A}$ and $\boldsymbol{A}'$ are $\boldsymbol{K}$-monadic (or synonymously, $\boldsymbol{K}$-tripleable) over $\boldsymbol{K}$, and that both $\boldsymbol{A}$ and $\boldsymbol{A}'$ are also $\boldsymbol{K}$-cocomplete. These facts will be exploited later on.

2.1. Proposition. *$\boldsymbol{A}$ and $\boldsymbol{A}'$ are cotensored $\boldsymbol{K}$-categories. Furthermore, the "underlying Kelley space" $\boldsymbol{K}$-functors preserve cotensors (strictly).*

Proof. The above statement just means that all the representable functors $\boldsymbol{A}^{\mathrm{op}} \xrightarrow{A(-,A)} \boldsymbol{K}$ have a $\boldsymbol{K}$-left adjoint. In order to prove it, it will be enough to show that for all $A \in \boldsymbol{A}$ and $X \in \boldsymbol{K}$, the cotensor of A with X exists, or in other words, there is an object $\bar{\boldsymbol{A}}(X, A) \in \boldsymbol{A}$ and a $\boldsymbol{K}$-natural isomorphism

$$\boldsymbol{A}(-, \bar{A}(X, A)) \overset{\sigma}{\approx} \boldsymbol{K}(X, \boldsymbol{A}(-, A)).$$

Define $|\bar{\boldsymbol{A}}(X, A)| = \boldsymbol{K}(X, |A|)$ with operations:

$$\boldsymbol{K}(X, |A|) \boxtimes \boldsymbol{K}(X, |A|) \approx \boldsymbol{K}(X, |A| \boxtimes |A|) \underset{\boldsymbol{K}(X, \text{"}\cdot\text{"})}{\overset{\boldsymbol{K}(X, \text{"}+\text{"})}{\rightrightarrows}} \boldsymbol{K}(X, |A|),$$

$$\boldsymbol{K}(X, |A|) \xrightarrow{\boldsymbol{K}(X, \text{"}-\text{"})} \boldsymbol{K}(X, |A|),$$

$$\frac{C \boxtimes \boldsymbol{K}(X, |A|) \longrightarrow \boldsymbol{K}(X|A|)}{\dfrac{C \to \boldsymbol{K}(|A|, |A|)}{C \boxtimes |A| \xrightarrow{\cdot} |A|}\ \omega_0 \xrightarrow{\boldsymbol{K}(X, -)} \boldsymbol{K}(\boldsymbol{K}(X, |A|), \boldsymbol{K}(X, |A|))}\ \omega_0 .$$

and
$$\frac{1 \to \boldsymbol{K}(X, |A|)}{1 \boxtimes X \to 1 \xrightarrow{0} |A|}\ \omega_0 .$$

A routine diagram decomposition process shows that these operations make all diagrams in Definition 1.1 commutative, and therefore $\bar{A}(X,A) \in \mathbf{A}$. It can be checked that the above definitions produce the standard point-wise operations on functions. The advantage of this presentation resides not in the algebraic properties to be checked, but in the fact that the continuity is automatically guaranteed.

Given any other $B \in \mathbf{A}$, consider the diagram

$$\begin{tikzcd} \mathbf{A}(B,\bar{A}(X,A)) \arrow[d, "|\ |"] & \mathbf{K}(X,\mathbf{A}(B,A)) \arrow[d, "\mathbf{K}(X,|\ |)"] \\ \mathbf{K}(|B|,\mathbf{K}(X,|A|)) \arrow[r, "\sigma", "\approx"'] & \mathbf{K}(X,\mathbf{K}(|B|,|A|)) \end{tikzcd} .$$

It is easy to see that the isomorphism σ restricted to the upper level provides a bijection. Since both vertical arrows are full injections (see 0.14), this bijection is bicontinuous, i.e., an isomorphism in $\mathbf{K}$. The $\mathbf{K}$-naturality follows now from the $\mathbf{K}$-naturality in the lower level and the fact that $|\ |$ is a $\mathbf{K}$-faithful $\mathbf{K}$-functor. Finally, the commutativity of the diagram above [completed with $\mathbf{A}(B,\bar{A}(X,A)) \to \mathbf{K}(X,\mathbf{A}(B,A))$] means that $|\ |$ preserves (strictly) the cotensor just constructed.

For $\mathbf{A}'$ we manipulate similarly: if $A \in \mathbf{A}'$, define $\bar{A}'(X,A) = \bar{A}(X,A)$ with the identity

$$\frac{1 \boxtimes X \to 1 \xrightarrow{e} |A|}{1 \to \mathbf{K}(X,|A|)} \quad \omega_0 .$$

The proof follows the same lines as in the case of $\mathbf{A}$. Q.E.D.

2.2. Proposition.

(a) *Given any functor* $\Gamma \xrightarrow{\Gamma} \mathbf{A}$ *such that the composite* $\Gamma \xrightarrow{\Gamma} \mathbf{A} \xrightarrow{|\ |} \mathbf{K}$ *has a limit* ($= \varprojlim$), *then* $\Gamma \xrightarrow{\Gamma} \mathbf{A}$ *has also a limit which is strictly preserved by* $\mathbf{A} \xrightarrow{|\ |} \mathbf{K}$. *Furthermore, given any* $A \in \mathbf{A}$, *the limit of* Γ *is also preserved under* $\mathbf{A} \xrightarrow{\mathbf{A}(A,-)} \mathbf{K}$.

(b) *Similar to* (a) *with* $\mathbf{A}$ *replaced by* $\mathbf{A}'$

Proof. Define $|\varprojlim \Gamma_\lambda| = \varprojlim |\Gamma_\lambda|$ with operations:

$$\begin{tikzcd} \varprojlim_\lambda |\Gamma_\lambda| \boxtimes \varprojlim_\lambda |\Gamma_\lambda| \arrow[r, dashed, shift left] \arrow[r, dashed, shift right] \arrow[d, dashed, "P_\lambda \boxtimes P_\lambda"] & \varprojlim_\lambda |\Gamma_\lambda| \arrow[d, dashed, "P_\lambda"] \\ |\Gamma_\lambda| \boxtimes |\Gamma_\lambda| \arrow[r, shift left, "+"] \arrow[r, shift right, "\cdot"'] & |\Gamma_\lambda| \end{tikzcd} , \quad \begin{tikzcd} \varprojlim_\lambda |\Gamma_\lambda| \arrow[r, dashed] \arrow[d, dashed, "P_\lambda"] & \varprojlim_\lambda |\Gamma_\lambda| \arrow[d, dashed, "P_\lambda"] \\ |\Gamma_\lambda| \arrow[r, "-"] & |\Gamma_\lambda| \end{tikzcd} ,$$

$$\begin{tikzcd} C \boxtimes \varprojlim_\lambda |\Gamma_\lambda| \arrow[r, dashed] \arrow[d, dashed, "\mathrm{id} \boxtimes P_\lambda"] & \varprojlim_\lambda |\Gamma_\lambda| \arrow[d, dashed, "P_\lambda"] \\ C \boxtimes |\Gamma_\lambda| \arrow[r, "\cdot"] & |\Gamma_\lambda| \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} 1 \arrow[rr, dashed] \arrow[dr, dashed, "0"'] & & \varprojlim_\lambda |\Gamma_\lambda| \arrow[dl, dashed, "P_\lambda"] \\ & |\Gamma_\lambda| & \end{tikzcd} .$$

The fact that for any $\lambda \xrightarrow{f} \mu$ in $\mathbf{\Gamma}$, $|\Gamma_\lambda| \xrightarrow{\Gamma(f)} |\Gamma_\lambda|$ is a morphism of K-topological algebras, is all what it is needed in order to check the commutativities required for the existence of the dotted arrows. It is indeed a straightforward matter that the axioms (Definition 1.1) are satisfied. In order to see that the resulting K-topological algebra is actually the $\varprojlim$ of Γ and that it is preserved by the representable functors, consider the diagram:

$$\begin{array}{ccc} \mathbf{A}(A, \varprojlim_\lambda \Gamma_\lambda) & & \varprojlim_\lambda \mathbf{A}(A, \Gamma_\lambda) \\ \downarrow{\scriptstyle |\ |} & & \downarrow{\scriptstyle \varprojlim_\lambda |\ |} \\ \mathbf{K}(|A|, \varprojlim_\lambda |\Gamma_\lambda|) & \approx & \varprojlim_\lambda \mathbf{K}(|A|, |\Gamma_\lambda|) \end{array} .$$

The homeomorphism at the lower level (which is in fact a homeomorphism because $\mathbf{K}(|A|, -)$ has a left adjoint) induces by restriction a bijection in the upper level which is continuous in both directions because both vertical arrows are full injections (it can be checked easily that a $\varprojlim$ of full injections is a full injection). Similar arguments apply to $\mathbf{A}'$. Q.E.D.

Since $\mathbf{K}$ is a complete category (see 0.7), it follows from the Proposition above that $\mathbf{A}$ and $\mathbf{A}'$ are also complete. The fact that the limits in $\mathbf{A}$ and $\mathbf{A}'$ are preserved by the representables into $\mathbf{K}$ means (see [10] for definitions) that they are ***K**-limits.* This together with Proposition 2.1 amount to saying that *$\mathbf{A}$ and $\mathbf{A}'$ are **K**-complete **K**-categories.* It also follows from the proofs of Prop. 2.1 and 2.2 that the inclusion $\mathbf{A}' \to \mathbf{A}$ is a $\mathbf{K}$-functor which preserves limits and cotensors.

We proceed to prove now some other facts promised at the beginning of this section.

Let X be a Kelley space and A a K-topological algebra. We say that *X generates A (via f)* if there is a continuous function $X \xrightarrow{f} |A|$ (i.e., a morphism in $\mathbf{K}$) such that the set-theoretical image of f algebraically generates A (or, no proper purely algebraic subalgebra of A contains the image of f). *The class of all K-topological algebras generated by any given Kelley space X is a set.* In fact, there is only a set of surjective functions with domain X. For each of them, there is only a set of algebras algebraically generated, and finally, for each of those there is only a set of possible topologies.

A similar definition and conclusion are clear in the case of K-topological algebras with identity.

2.3. Proposition. *The $\mathbf{K}$-functors $\mathbf{A} \xrightarrow{|\ |} \mathbf{K}$ and $\mathbf{A}' \xrightarrow{|\ |} \mathbf{K}$ have left adjoints $\mathbf{K} \xrightarrow{F} \mathbf{A}$ and $\mathbf{K} \xrightarrow{F'} \mathbf{A}'$. Furthermore, F and F' are **K**-functors and **K**-left adjoints.*

Proof. Since $\mathbf{A}$ is well-powered and $\mathbf{A} \xrightarrow{|\ |} \mathbf{K}$ preserves limits, by the Adjoint Functor Theorem [15] it is enough to obtain, for any given $X \in \mathbf{K}$, a solution set. But the set

of K-topological algebras generated by X furnishes a solution. In fact, let $X \xrightarrow{g} |A|$ be any map in $\boldsymbol{K}$ and let $I \in \boldsymbol{A}$ be the algebraic subalgebra of A generated by the set-theoretical image of g endowed with the K-ation of the relative topology corresponding to $I \subset A$. It is clear that I is a K-topological algebra and the inclusion $I \xrightarrow{i} A$ is a morphism in $\boldsymbol{A}$. The map $X \xrightarrow{g} |A|$ has a factorization

$$\begin{tikzcd} X \arrow[r, dashed, "g"] \arrow[dr, dashed, "f"'] & {|A|} \\ & {|I|} \arrow[u, dashed, "i"'] \end{tikzcd}$$

Since i is a full injection, f is continuous and it is clear that X generates I via f, therefore the set of K-topological algebras generated by X is a solution set, as claimed. Thus, $\boldsymbol{A} \xrightarrow{|\,|} \boldsymbol{K}$ has a left adjoint $\boldsymbol{K} \xrightarrow{F} \boldsymbol{A}$. Since $\boldsymbol{A}$ is cotensored and $|\;|$ preserves cotensors, the last part of 2.3 follows as an application of the criterion given in [23] 4.1, p. 173. The corresponding results for $\boldsymbol{A}' \xrightarrow{|\,|} \boldsymbol{K}$ are obtained in the same way.

Q.E.D.

A description of FX, $X \in \boldsymbol{K}$ can be given as follows: let $V(X)$ be the free complex vector space over X (which can be pictured as the space of all functions a: $X \to C$ such that $a(x) \neq 0$ holds only for finitely many $x \in X$). A topology on $V(X)$ is determined by the convergence $a_\alpha \to a$ if and only if for each $A \in \boldsymbol{A}$ and $X \xrightarrow{\varphi} |A|$, φ a morphism in $\boldsymbol{K}$, we have $\Sigma\{a_\alpha(x)\varphi(x); x \in X\} \to \Sigma\{a(x)\varphi(x), x \in X\}$. This topology can be lifted to the tensor algebra $T[V(X)] = V(X) \oplus (V(X) \otimes_C V(X)) \oplus \ldots$ and $F(X)$ = largest Hausdorff quotient of $\mathrm{K}e T[V(X)]$. Similarly, $F'A$ is an extension of FX by C with trivial action (cf. Proposition 2.10).

2.4. Proposition. *The K-functors $\boldsymbol{A} \xrightarrow{|\,|} \boldsymbol{K}$ and $\boldsymbol{A}' \xrightarrow{|\,|} \boldsymbol{K}$ are (strictly) K-monadic. More specifically, $\boldsymbol{A}$ and $\boldsymbol{A}'$ are (K-isomorphic to) the K-categories of algebras over the K-monads determined in K by the pairs of K-adjoint functors $F \dashv_{\boldsymbol{K}} |\;|$ and $F' \dashv_{\boldsymbol{K}} |\;|$.*

Proof. This result is an easy application of the enriched version of Beck's Tripleability Theorem (cf. [10] Theorem II. 2.1). There is no difficulty in checking the hypotheses for the "underlying Kelley space" $\boldsymbol{K}$-functors $\boldsymbol{A} \xrightarrow{|\,|} \boldsymbol{K}$ and $\boldsymbol{A}' \xrightarrow{|\,|} \boldsymbol{K}$. Q.E.D.

2.5. Remark. *Given any map $A \xrightarrow{\varphi} B$ in $\boldsymbol{A}$ (resp. $\boldsymbol{A}'$) φ can be factored in $\boldsymbol{A}$ (resp. $\boldsymbol{A}'$)*

$$\begin{tikzcd} A \arrow[r, "\varphi"] \arrow[dr, dashed, "\psi"'] & B \\ & I \arrow[u, dashed, "i"'] \end{tikzcd}$$

where ψ is a surjective function and i is a full injection.

Proof. Define I to be the set-theoretic image of φ with the K-ation of the relative topology. It is easy to check that $I \in \boldsymbol{A}$ and clearly φ factors as $A \xrightarrow{\psi} I \xrightarrow{i} B$. Finally, ψ is continuous because i is a full injection. Q.E.D.

2.6. Proposition. *The $\boldsymbol{K}$-categories $\boldsymbol{A}$ and $\boldsymbol{A}'$ have all coequalizers.*

Proof. Let $A \xrightarrow{\varphi} B$, $A \xrightarrow{\psi} B$ be any pair of maps in $\boldsymbol{A}$. From 2.5 follows that there is a solution set for the coequalizer of φ and ψ (namely the set of all $B \xrightarrow{d} D$ in $\boldsymbol{A}$ which are surjective functions and such that $d\varphi = d\psi$) and therefore the coequalizer of φ and ψ, $B \xrightarrow{h} H$ does exist. In fact, form the category $\boldsymbol{\Gamma}$ whose objects are maps $B \xrightarrow{d} D$ as above and whose arrows $d \xrightarrow{l} d'$ are maps $D \xrightarrow{l} D'$ in $\boldsymbol{A}$ such that $d' = ld$. $\boldsymbol{\Gamma}$ is a small category and there is a functor $\boldsymbol{\Gamma} \xrightarrow{\Gamma} \boldsymbol{A}$, $\Gamma(B \xrightarrow{d} D) = D$, $\Gamma l = l$. Since $\boldsymbol{A}$ is complete, the (inverse) limit of Γ exists. Thus, H is this limit with $B \xrightarrow{h} H$ defined as follows

$$\begin{tikzcd} B \arrow[rr, "\exists ! h"] \arrow[dr, "d"'] & & H \arrow[dl, "P_d"] \\ & D & \end{tikzcd}$$

where P_d is the projection corresponding to $(B \xrightarrow{d} D) \in \boldsymbol{\Gamma}$. The same is done in the case of $\boldsymbol{A}'$. Q.E.D.

Since $\boldsymbol{K}$ is a cocomplete category (see 0.8) and the $\boldsymbol{K}$-functors $\boldsymbol{A} \xrightarrow{|\,|} \boldsymbol{K}$ and $\boldsymbol{A}' \xrightarrow{|\,|} \boldsymbol{K}$ are $\boldsymbol{K}$-monadic, and in particular, monadic (= tripleable), it follows from Prop. 2.6 and a well known result of Linton [24] that $\boldsymbol{A}$ and $\boldsymbol{A}'$ have all (small) colimits. We state this:

2.7. Proposition. *The $\boldsymbol{K}$-categories $\boldsymbol{A}$ and $\boldsymbol{A}'$ have all (small) colimits. Furthermore, they are preserved by the (contravariant) representables into $\boldsymbol{K}$ and therefore they are $\boldsymbol{K}$-colimits.*

Proof. It only remains to be seen that the representables $\boldsymbol{A}^{\mathrm{op}} \xrightarrow{\boldsymbol{A}(-,A)} \boldsymbol{K}$ preserve colimits. But this is clear since $\boldsymbol{A}$ is cotensored and therefore the $\boldsymbol{K}$-functors $\boldsymbol{A}(-,A)$, have left adjoints. Q.E.D.

The statement above reads: given any functor $\boldsymbol{\Gamma} \xrightarrow{\Gamma} \boldsymbol{A}$ where $\boldsymbol{\Gamma}$ is small, then $\underset{\overrightarrow{\lambda}}{\operatorname{colim}}\, \Gamma_\lambda$ exists in $\boldsymbol{A}$ and for any K-topological algebra $A \in \boldsymbol{A}$ there is a homeomorphism of the Kelley spaces:

$$\boldsymbol{A}(\underset{\overrightarrow{\lambda}}{\operatorname{colim}}\, \Gamma_\lambda, A) \approx \underset{\overleftarrow{\lambda}}{\lim}\, \boldsymbol{A}(\Gamma_\lambda, A)\,.$$

Let us observe now that, as in the case of $\boldsymbol{K}$ (see 0.14), the representables of $\boldsymbol{A}$ or any $\boldsymbol{K}$-category of K-topological algebras (Prop. 1.7) preserve full injections. More precisely:

2.8. Remark. Given $I \xrightarrow{i} B$ in $\boldsymbol{A}$ such that $|I| \xrightarrow{i} |B|$ is a full injection (in $\boldsymbol{K}$), then for all $A \in \boldsymbol{A}$, $\boldsymbol{A}(A, I) \xrightarrow{\boldsymbol{A}(A,i)} \boldsymbol{A}(A, B)$ is also a full injection (in $\boldsymbol{K}$).

Proof. In the diagram

$$\begin{tikzcd}
\boldsymbol{A}(A,I) \arrow[r, "\boldsymbol{A}(A,i)"] \arrow[d, "|\ |"'] & \boldsymbol{A}(A,B) \arrow[d, "|\ |"] \\
\boldsymbol{K}(|A|,|I|) \arrow[r, "\boldsymbol{K}(|A|,i)"] & \boldsymbol{K}(|A|,|B|)
\end{tikzcd}$$

the two vertical arrows and the lower level arrow are full injections, whence the upper level arrow is also a full injection. Q.E.D.

2.9. Proposition. *$\boldsymbol{A}$ and $\boldsymbol{A}'$ are tensored $\boldsymbol{K}$-categories.*

Proof. The meaning of this statement is that all representables $\boldsymbol{A} \xrightarrow{\boldsymbol{A}(A,-)} \boldsymbol{K}$ have a $\boldsymbol{K}$-left adjoint. But these functors preserve limits (Prop. 2.2) and $\boldsymbol{A}$ is well powered (and complete), so that, by the Adjoint Functor Theorem there will exist left adjoints provided that for any given $X \in \boldsymbol{K}$ there is a solution set. Let $A \in \boldsymbol{A}$ and $X \in \boldsymbol{K}$ be fixed objects. Denote by $A \xrightarrow{i_x} \coprod_X A$ the coproduct in $\boldsymbol{A}$ of A repeated as a factor once for each point of X. Given any $B \in \boldsymbol{A}$ and $X \xrightarrow{f} \boldsymbol{A}(A, B)$ in $\boldsymbol{K}$, let $\coprod_X A \xrightarrow{\omega_0(f)} B$ be the map (in $\boldsymbol{A}$) defined by the diagram

$$\begin{tikzcd}
\coprod_X A \arrow[rr, dashed, "\exists !"] & & B \\
& A \arrow[ul, "i_x"] \arrow[ur, "f(x)"'] &
\end{tikzcd} \qquad x \in X.$$

It is clear that the correspondence ω_0 is one-to-one. Let now $S = \{ X \xrightarrow{h} \boldsymbol{A}(A, H);\ H \in \boldsymbol{A}, h \in \boldsymbol{K}, \omega_0(h) \text{ is onto} \}$. Since there is only a set of surjective functions $\coprod_X A \to H$, S is a set. In order to see that S is a solution set we proceed as follows. Let $X \xrightarrow{f} \boldsymbol{A}(A, B)$ and consider the factorization of $\omega_0(f)$ described in Remark 2.5:

$$\begin{tikzcd}
\coprod_X A \arrow[rr, "\omega_0(f)"] \arrow[dr, "g"'] & & B \\
& H \arrow[ur, "i"'] &
\end{tikzcd}$$

Define $X \xrightarrow{h} \boldsymbol{A}(A, H)$ by $h(x) = (A \xrightarrow{i_x} \coprod_X A \xrightarrow{g} H)$. The diagram

$$\begin{tikzcd} X \arrow[r, dashed, "f"] \arrow[d, dashed, "h"'] & A(A,B) \\ A(A,H) \arrow[ur, "A(A,i)"'] & \end{tikzcd}$$

clearly commutes, and therefore, from Remark 2.8 follows that h is continuous, or $h \in \boldsymbol{K}$. Finally, it is obvious that $\omega_0(h)\, i_x = g i_x$ (for all $x \in X$) and therefore $\omega_0(h) = g$. Hence $\omega_0(h)$ is onto, that is, $h \in S$. This completes the proof of the existence of a left adjoint for $\boldsymbol{A}(A, -)$. Since the representables preserve cotensors (always), it follows as in Prop. 2.3 that this left adjoint is necessarily a $\boldsymbol{K}$-functor and a $\boldsymbol{K}$-left adjoint. This proof can be adapted without difficulty to $\boldsymbol{A}'$. Q.E.D.

We finish this section by expressing $\boldsymbol{K}$-functorially the standard procedure of adding an identity to an algebra possibly lacking it.

2.10. Proposition. *The $\boldsymbol{K}$-inclusion $\boldsymbol{A}' \to \boldsymbol{A}$ has a $\boldsymbol{K}$-left adjoint $\boldsymbol{A} \xrightarrow{(\widetilde{-})} \boldsymbol{A}'$. That is, given any K-topological algebra A there is a K-topological algebra with identity $\tilde{A}$ and a natural homeomorphism $\boldsymbol{A}(A,B) \approx \boldsymbol{A}'(\tilde{A},B)$ (for all K-topological algebras with identity B). At the level of sets, we have a bijection*

$$\frac{A \longrightarrow B}{\tilde{A} \longrightarrow B}$$

between continuous linear multiplicative functions $A \to B$ and continuous linear multiplicative functions $\tilde{A} \to B$ which preserve the identity.

Proof. This result follows, for example, from Prop. 2.4 and 2.6 and Theorem A.1 in the Appendix of [10]. It is only necessary to observe that the inclusion $\boldsymbol{A}' \to \boldsymbol{A}$ commutes with the "underlying Kelley space" $\boldsymbol{K}$-functors.

§ 3. Gelfand $\boldsymbol{K}$-monads, the duality determined by C

Given a Kelley space $X \in \boldsymbol{K}$, consider the cotensor $\bar{\boldsymbol{A}}(X, C) = \bar{\boldsymbol{A}}'(X, C)$ (equality occurs since C belongs to both $\boldsymbol{A}$ and $\boldsymbol{A}'$ and $\boldsymbol{A}' \xrightarrow{i} \boldsymbol{A}$ preserves cotensors). According to the definitions given (§2) this is nothing but the long-considered algebra of all complex valued functions on X, endowed with the K-ation of the compact-open topology, in other words, the K-ation of the topology of uniform convergence on compact sets of X. In particular, $\bar{\boldsymbol{A}}(X, C)$ is the K-ation of a complete locally m-convex algebra (cf. [27], Appendix D). If K is a compact space, then $\bar{\boldsymbol{A}}(K, C)$ is just the commutative C^*-algebra of all continuous complex valued functions on K with the supremum norm, which is already a k-space (0.1). Since the $\boldsymbol{K}$-functor

$\mathbf{K}^{\mathrm{op}} \xrightarrow{A(-,C)} \mathbf{A}$ preserves limits, we have $A(X,C) = \varprojlim_{K \subset X} A(K,C)$ for all $X \in \mathbf{K}$ (where K stands for an arbitrary compact subset of X). Thus, $A(X,C)$ is always a filtered inverse limit of commutative C^*-algebras (recall however that this limit is taken qua K-topological algebras). These limits are easily characterized; we write now the "K-ation" of a well known result:

3.1. Proposition. *Given a K-topological algebra $A \in \mathbf{A}$, the following are equivalent:*

(a) *A is a limit (in $\mathbf{A}$) of commutative C^*-algebras with identity;*

(b) *A is commutative with identity and there is a family of algebra seminorms $\{p\}$ which defines the topology of A (in the sense that $|A|$ is the K-ation of the locally m-convex algebra $(A, \{p\})$). Furthermore the locally m-convex algebra $(A, \{p\})$ is complete and there is an involution $*$ satisfying $p(a^*a) = p(a)^2$ for all $a \in A$ and all p.*

Proof. (a) ⇒ (b) is straightforward. For (b) ⇒ (a), observe that $A = \varprojlim_p A_p$ where A_p is the completion of the quotient of A by the null set of p; A_p is a C^*-algebra and $(A, \{p\}) = \varprojlim_p A_p$ (lim in the category of locally convex algebras). For more details see [17], [19], [27]. Q.E.D.

We remark that in the same fashion, the equivalence of the following statements about a K-topological algebra A can be established:

(a′) *A is a limit (in $\mathbf{A}$) of Banach algebras;*

(b′) *there is a family $\{p\}$ of algebra seminorms which defines the topology of A (in the above sense) and such that the locally m-convex algebra $(A, \{p\})$ is complete.* As above, we have now $A = \varprojlim A_p$.

Let us remark that given an algebra over the complex numbers (in the purely algebraic sense) non-equivalent families of algebra seminorms may determine the same K-topological algebra (by non-equivalent families of seminorms we mean to understand that the induced locally convex topologies do not coincide). For example, given a locally m-convex algebra A whose topology is determined by a family $\{p\}$ of seminorms, we can enlarge $\{p\}$ by adding in any set of (in particular, all) seminorms continuous on compact subsets of A. It is clear that the locally convex topologies may disagree, and yet, the K-topological algebras determined by these two families necessarily coincide. Also, starting from an $A \in \mathbf{A}$ determined by a *complete* locally m-convex algebra, the process of adding seminorms $\{q\}$ just described leads to a new locally m-convex algebra B that may not be complete, in which case A, although still determined by B may not coincide with the enlarged limit $\varprojlim B_q$. We describe now some of these phenomena.

3.2. Example. Let X be the locally compact space of ordinals $X = [1,\Omega)$ with the

order topology, where Ω is the smallest uncountable ordinal, A^{b} the locally m-convex algebra of all continuous complex valued functions on X with the topology of uniform convergence on compact sets. We have $\bar{A}(X,C) = \mathrm{Ke}A^{\mathrm{b}}$. It is well known that every function $f \in A^{\mathrm{b}}$ is constant on a tail $[\alpha, \Omega)$ for some α (which depends of course on f) and therefore, f has an extension f^0 to $\beta X = [1, \Omega]$. The correspondence $f \rightsquigarrow f^0$ is an algebraic isomorphism (onto) between A^{b} and $\bar{A}(\beta X, C)$. Denote now by $L \in A(\bar{A}(\beta X, C), C)$ the continuous multiplicative linear map $L(f) = f(\Omega)$. It is clear that L induces a multiplicative linear map (also denoted by L) from A^{b} to C.

3.2.1. *$L \in A(\bar{A}(X, C), C)$, but L is not continuous on A^{b}.*

Proof. First, for each $\alpha < \Omega$ denote by p_α the seminorm $p_\alpha(f) = \mathrm{Sup}\{|f(\gamma)|;\ 1 \leqslant \gamma \leqslant \alpha\}$. It is easy to see that the family $\{p_\alpha\}$, $1 \leqslant \alpha < \Omega$ determines the topology of A^{b}. Now, let $g_\alpha \in A^{\mathrm{b}}$ be the function defined by $g_\alpha(\gamma) = 0$ if $1 \leqslant \gamma \leqslant \alpha$, $g_\alpha(\gamma) = 1$ if $\gamma \geqslant \alpha + 1$, where $1 \leqslant \alpha < \Omega$. It is clear that $p_\alpha(g_\sigma) = 0$ if $\sigma > \alpha$, and therefore $g_\sigma \to 0$ (as $\sigma \to \Omega$) in A^{b} (0 is the zero function). However, $L(g_\sigma) = g_\sigma^0(\Omega) = 1$ which shows that L *is not continuous* on A^{b}. In order to prove the first part, we observe that $L \in A(\bar{A}(X, C), C)$ just means that the restrictions of L to the compact subsets of A^{b} are continuous. This follows from the fact that if $H \subset A^{\mathrm{b}}$ is a compact subset, then there is an ordinal $\delta < \Omega$ such that *all* functions $f \in H$ are constant on $[\delta, \Omega)$ ([4] Chap. IV, §4, Ex. 17). For the sake of completeness, we sketch here the proof. First, there is a constant $M > 0$ such that $|f(\sigma)| \leqslant M$ for all $f \in H$ and $\sigma \in X$, since otherwise there would exist ordinals $\sigma_n < \Omega$, $n = 1, 2, \ldots$ and functions $f_n \in H$ with $|f_n(\sigma_n)| \geqslant n$; but necessarily $\sigma = \mathrm{Sup}\,\sigma_n < \Omega$ and $p_\sigma(f)$ is bounded on H (compact), which contradicts the above conclusion. Now define for $\sigma \in X$ the number $s(\sigma) = \mathrm{Sup}\{|f(\sigma') - f(\sigma'')|;\ f \in H,\ \sigma \leqslant \sigma',\ \sigma \leqslant \sigma''\}$. Clearly $s(\sigma) \leqslant 2M$ and $\sigma(s)$ decreases as $\sigma \to \Omega$; this means that s is eventually constant: $s(\sigma) = u$ for all $\sigma \geqslant \sigma_0$ in X. If $u > 0$, there exist sequences $\{\sigma_n\}$ and $\{\tau_n\}$ in X and $\{f_n\}$ in H such that $\sigma_n \leqslant \tau_n \leqslant \sigma_{n+1} \leqslant \tau_{n+1}$, $n = 1, 2, \ldots$ and $|f_n(\sigma_n) - f_n(\tau_n)| \geqslant \frac{1}{2}u$; clearly both $\{\sigma_n\}$ and $\{\tau_n\}$ converge to $\bar{\sigma} = \mathrm{Sup}\{\sigma_n\} = \mathrm{Sup}\{\tau_n\}$, and $\{f_n\}$ has a convergent subset $\{f_{n'}\}$ with limit $f \in H$. This leads to the contradiction $f(\bar{\sigma}) = \lim f_{n'}(\sigma_n) \neq \lim f_{n'}(\tau_n) = f(\bar{\sigma})$. Thus, $u = 0$ and $f(\sigma') = f(\sigma'')$ for all $f \in H$ and $\sigma' \geqslant \sigma_0$, $\sigma'' \geqslant \sigma_0$, as desired. From here it follows trivially that the restriction of L to H is continuous, as claimed. Q.E.D.

Observe that $q_L(f) = |L(f)|$ is an algebra seminorm on A^{b}, continuous on compact subsets. Thus the locally m-convex algebra B defined by the family $\{p_\sigma\} \cup \{q_L\}$ on A^{b} satisfies $\mathrm{Ke}A^{\mathrm{b}} = \mathrm{Ke}B$. However A^{b} (with $\{p_\sigma\}$) is complete, and B (with $\{p_\sigma\}, q_L$) *is not:* the net $\{g_\sigma\}_{1 \leqslant \sigma < \Omega}$ defined above is a Cauchy net for p_σ and q_L, but does not converge in B; as is easily seen. This shows in particular that $B \neq A^{\mathrm{b}}$ and therefore all the statements made in the last paragraph before 3.2 have now been justified. This example will be continued later (see 3.6 below).

Given a K-topological algebra $A \in \boldsymbol{A}$, $\boldsymbol{A}(A, C)$ is the set of all continuous linear multiplicative functionals on A with the K-ation of its compact-open topology. If $A \in \boldsymbol{A}'$, then $\boldsymbol{A}'(A, C)$ consists only of those functionals that preserve the identity. In the case where A is determined by (i.e., is obtained as the K-ation of) a locally m-convex algebra B, the above spaces differ from what has been classically called the "spectrum" or "carrier space" of B ([5], [9], [17]) in two different ways. On one hand, they contain more points (namely, all those characters [= linear multiplicative functionals] which are continuous on compact sets, but not continuous on B); on the other hand, the topology of $\boldsymbol{A}(A, C)$ (and $\boldsymbol{A}'(A, C)$) is *not* the customary topology for spectra, namely, the topology of simple convergence on the elements of B (or even the K-ation of it). These facts support a statement asserting that $\boldsymbol{A}(A, C)$ and $\boldsymbol{A}'(A, C)$ differ substantially from the traditionally considered spectra. And yet, if A $(= B)$ is a commutative C^*-algebra with identity, then $\boldsymbol{A}'(A, C)$ coincides with the spectrum of A, set-wise and even topologically. In fact, the only thing to be verified is that pointwise convergence of characters coincides with uniform convergence on compact sets, an obvious fact since in this case the characters are equicontinuous. In particular $\boldsymbol{A}'(A, C)$ is a compact space; $\boldsymbol{A}(A, C)$ is also compact and in fact is obtained from $\boldsymbol{A}'(A, C)$ by adding one isolated point: the zero functional. It follows from these considerations that $\boldsymbol{A}'(A, C)$ is a legitimate generalization of the spectrum of A when A is C^*, and therefore $\boldsymbol{A}'(A, C)$, for $A \in \boldsymbol{A}$ should play a similar role than the spectrum concerning, for instance, the existence of idempotents, representations, etc., and in particular the Gelfand Theory of C^*-algebras. The rest of the section is devoted to developing a generalization of the latter.

Consider the K-category $\boldsymbol{A}$ and the pair of $\boldsymbol{K}$-adjoint functors

$$\boldsymbol{A} \underset{\bar{\boldsymbol{A}}(-, C)}{\overset{\boldsymbol{A}(-, C)}{\rightleftarrows}} \boldsymbol{K}^{\mathrm{op}} .$$

The corresponding $\boldsymbol{K}$-monad in $\boldsymbol{A}$ will be called the *Gelfand* $\boldsymbol{K}$*-monad*, denoted $\boldsymbol{T} = (T, \eta, \mu)$, where $\boldsymbol{A} \xrightarrow{T} \boldsymbol{A}$, $TA = \bar{\boldsymbol{A}}(\boldsymbol{A}(A, C), C)$ for any given K-topological algebra $A \in \boldsymbol{A}$. The unit $\mathrm{id} \overset{\eta}{\Rightarrow} T$ is the *Gelfand transformation* (sometimes called *Fourier transformation*, or even *Fourier-Gelfand transformation*): $A \xrightarrow{\eta A} \bar{\boldsymbol{A}}(\boldsymbol{A}(A, C), C)$. If $a \in A$, we introduce the notation $\eta A(a) = \hat{a}$, where $\hat{a}(\psi) = \psi(a)$ for any $A \xrightarrow{\psi} C$ in $\boldsymbol{A}$. The multiplication of the monad $TT \overset{\mu}{\Rightarrow} T$, $\bar{\boldsymbol{A}}(\boldsymbol{A}(\bar{\boldsymbol{A}}(\boldsymbol{A}(A, C), C), C), C) \xrightarrow{\mu A} \bar{\boldsymbol{A}}(\boldsymbol{A}(A, C), C)$ has the following action: given $\boldsymbol{A}(\bar{\boldsymbol{A}}(\boldsymbol{A}(A, C), C), C) \xrightarrow{\varphi} C$ and $A \xrightarrow{\psi} C$, then $\mu A(\varphi)(\psi) = \varphi(\dot{\psi})$ where $\bar{\boldsymbol{A}}(\boldsymbol{A}(A, C), C) \xrightarrow{\dot{\psi}} C$ is defined on $\boldsymbol{A}(A, C) \xrightarrow{f} C$ by $\dot{\psi}(f) = f(\psi)$. (Recall that, as for any $\boldsymbol{K}$-monad determined by a pair of $\boldsymbol{K}$-adjoint functors, $\mu A = \bar{\boldsymbol{A}}(\epsilon \boldsymbol{A}(A, C), C)$, where ϵ is the counit (in $\boldsymbol{K}^{\mathrm{op}}$). Given $X \in \boldsymbol{K}$, $x \in X$, then $X \xrightarrow{\epsilon X} \boldsymbol{A}(\bar{\boldsymbol{A}}(X, C), C)$ is $\epsilon X(x) = \hat{x}$, where $\hat{x}(f) = f(x)$ for any $X \xrightarrow{f} C$ in $\boldsymbol{K}$.) The reader can verify easily that this description of the unit (Gelfand transformation) and multiplication of the Gelfand $\boldsymbol{K}$-monad actually describes the unit and multiplication as intrinsically obtained from the definition of the monad: it suffices to go back to the cotensoring isomorphism σ of Prop. 2.1, and observe that it is a

restriction of the isomorphism σ provided by the *closed Cartesian structure of* $\boldsymbol{K}$ (see 0.13). The continuity of the Gelfand transformation is automatically guaranteed, since it is a map in $\boldsymbol{A}$. Analogously, to the pair

$$\boldsymbol{A}' \underset{\bar{A}'(-,C)}{\overset{A'(-,C)}{\longleftarrow\!\!\longrightarrow}} \boldsymbol{K}^{\mathrm{op}}$$

there corresponds a $\boldsymbol{K}$-monad in $\boldsymbol{A}'$, that we also call the Gelfand $\boldsymbol{K}$-monad, and denote by $\boldsymbol{T}' = (T', \eta', \mu')$.

In order to indicate the overall picture, we take up space to review some past examples.

3.3. Example. If A is a commutative C^*-algebra with identity, then $T'A = A$. We will consider this in further detail below (see Prop. 3.7).

3.4. Example. Assume $X \in \boldsymbol{K}$ is completely regular (see [21]) and σ-compact, that is, X has a countable covering by compact subsets. Then, the locally m-convex algebra A^{b} of all continuous complex functions with the compact open topology is metrizable, and therefore $\bar{A}(X,C) = A^{\mathrm{b}}$. Assume now that $\varphi \in A(A^{\mathrm{b}}, C)$. Clearly for some compact subset $K \subset X$ we have $|\varphi(f)| \leqslant cp_K(f)$ where $p_K(f) = \operatorname{Sup}\{|f(k)|;\ k \in K\}$. Let $R \subset \bar{A}(K,C)$ be the algebra of restrictions to K of functions in A^{b} (i.e., R is the image of $\bar{A}(X,C) \xrightarrow{\bar{A}(i,C)} \bar{A}(K,C)$ where $K \xrightarrow{i} X$ is the inclusion. φ induces an element of $A(R,C)$. Let $J \subset R$ be the kernel of ψ: $g \in J$ if and only if $\psi(g) = 0$. If for every $x \in K$ there is a $g_x \in J$ such that $g_x(x) \neq 0$, then $h_x = |g_x|^2 = \bar{g}_x g_x \in J$ and h_x is real valued and satisfies $h_x(x) > 0$. A compactness argument shows that we can find $x_1, \ldots, x_n \in K$ with $h = \Sigma h_{x_i} > 0$ everywhere on K. But this means that h has an inverse in R and yet $\psi(h) = 0$. Thus for some $x \in K$ we have $g(x) = 0$ for all $g \in J$. Since J is maximal, the converse follows, and therefore for arbitrary g, $\varphi(g - g(x)1) = 0$, so that $\varphi(g) = g(x)$. This proves that $A'(A^{\mathrm{b}}, C)$ can be identified at the level of sets to X. However, if X is completely regular (and $X \in \boldsymbol{K}$), the topologies also agree and therefore $A'(A^{\mathrm{b}}, C) = X$. It follows that for these X, $T'\bar{A}(X,C) = \bar{A}(X,C)$. Observe that this equality follows from the immediate result $\bar{A}(X,C) = A^{\mathrm{b}}$ and Cor. 3.10. However, we have proved something stronger, namely that $X = A'(\bar{A}(X,C), C)$ as topological spaces.

3.5. Example. Let $\boldsymbol{S}$ be a set and $A = \mathrm{Ke}\,l^{\infty}(\boldsymbol{S})_{\mathrm{w}}$ (notation as in Ex. 1.3). From 1.3.4 follows easily that the family of functions in A with finite support (i.e., vanishing off a finite set) are dense in A. Then, if $\varphi \in A'(A,C)$ and φ is not zero, we must have $\varphi(a_s) \neq 0$ for some $s \in \boldsymbol{S}$, where $a_s \in A$ is the function $a_s(t) = 0$ if $t \neq s$, $a_s(s) = 1$. Since $a_s^2 = a_s$, clearly $\varphi(a_s) = 1$. If $a \in A$ is any element, we have $a_s(a - a(s)1) = 0$ so that $\varphi(a - a(s)1) = 1$, and therefore $\varphi(a) = a(s)$. This shows that $A'(A,C) = \boldsymbol{S}$ whence $T'A = T'\,\mathrm{Ke}\,l^{\infty}(\boldsymbol{S})_{\mathrm{w}} = \Pi_{\boldsymbol{S}}\, C \neq A$ (in $\boldsymbol{A}$). In fact, $T'T'A = T'A$. This follows, in case of $\boldsymbol{S}$ countable, from Ex. 3.4 above. In general it can be seen that $\prod_S C$ satisfies the conditions of Prop. 3.9 (see [31] Th. 5.2) and therefore Cor. 3.10 applies.

3.6. Example. This is a continuation of Example 3.2. X denotes again the space $X = [1, \Omega) \in \boldsymbol{K}$. Since $A = \boldsymbol{A}(X, C)$ is algebraically isomorphic (via $f \rightsquigarrow f^0$) to the C^*-algebra $\boldsymbol{A}(\beta X, C)$, it follows that the linear multiplicative functionals φ on A are the following: $\varphi = \varphi_\sigma$ defined as $\varphi_\sigma(f) = f(\sigma)$ for some $\sigma \in X$, which are obviously continuous, or $\varphi = L$, which according to 3.2.1 is also continuous. Thus $\boldsymbol{A}'(\boldsymbol{A}(X, C), C) = \beta X$. Since βX is compact, $B = \boldsymbol{A}(\beta X, C)$ is a C^*-algebra, and therefore $\boldsymbol{A}(B, C) = \beta X$ again. In other words, we have $T'T'\boldsymbol{A}'([1, \Omega), C) = T'\boldsymbol{A}'([1, \Omega), C) = \boldsymbol{A}'([1, \Omega], C) = \boldsymbol{A}'([1, \Omega), C)$. This last equality follows since both sides are algebraically isomorphic (via $f \rightsquigarrow f^0$), and this forces $\eta' A$ and $\boldsymbol{A}'(\epsilon X, C)$ to be mutually inverses of each other. We have then that $\mathrm{Ke}A^b$ is the C^*-algebra $\boldsymbol{A}'([1, \Omega], C)$, which indicates how substantially the K-ation functor changes the topology of a locally m-convex algebra. To add the semi-norm q_l is an intermediate step, that, although it makes every functional which is continuous over compact subsets be continuous, still does not constitute a Kelley topology. (Cf. [31] corollaries to Th. 5.2 for similar phenomena.)

Observe that the Gelfand $\boldsymbol{K}$-monad is the *codensity $\boldsymbol{K}$-monad* of the $\boldsymbol{K}$-functor $\boldsymbol{I} \xrightarrow{C} \boldsymbol{A}$ (cf. [10] p. 83), where $\boldsymbol{I}$ is the $\boldsymbol{K}$-category consisting of one single object $1 \in \boldsymbol{I}$ and $\boldsymbol{I}(1,1) = 1 \in \boldsymbol{K}$. A $\boldsymbol{K}$-functor $\boldsymbol{I} \to \boldsymbol{A}$ is characterized completely by one object of $\boldsymbol{A}$, and vice versa. This holds, of course, for all $\boldsymbol{K}$-categories, and not only for $\boldsymbol{A}$. Similarly, the Gelfand $\boldsymbol{K}$-monad in $\boldsymbol{A}'$ is the codensity $\boldsymbol{K}$-monad of $\boldsymbol{I} \xrightarrow{C} \boldsymbol{A}'$

Let now $\boldsymbol{C}^*$ denote the $\boldsymbol{K}$-category of commutative C^*-algebras with identity (see Prop. 1.2). Of course, $C \in \boldsymbol{C}^*$. From the remarks made immediately after Example 3.2, we know that for any $A \in \boldsymbol{C}^*$, $\boldsymbol{C}^*(A, C)$ is a compact space. On the other hand, given any compact space $K \in \boldsymbol{K}$, $\boldsymbol{A}'(K, C)$ is a commutative C^*-algebra with identity, and therefore, since $\boldsymbol{C}^* \xrightarrow{i} \boldsymbol{A}'$ is a $\boldsymbol{K}$-full subcategory, $\boldsymbol{C}^*(K, C) \overset{\text{def}}{=} \boldsymbol{A}'(K, C)$ is a cotensor of C with K in $\boldsymbol{C}^*$, so that $\boldsymbol{C}^*$ has, at least, cotensors of C with compact spaces. It follows that the codensity $\boldsymbol{K}$-monad of $\boldsymbol{I} \xrightarrow{C} \boldsymbol{C}^*$ exists, which simply means that we also have a Gelfand $\boldsymbol{K}$ monad in $\boldsymbol{C}^*$. It is clear that it is the restriction of $\boldsymbol{T}'$ to $\boldsymbol{C}^*$. The classical Gelfand duality (cf. for instance [5] or [28]) says that this $\boldsymbol{K}$-monad is isomorphic to the identity $\boldsymbol{C}^* \xrightarrow{\mathrm{id}} \boldsymbol{C}^*$ (recall that our $\boldsymbol{A}(A, C)$ coincides with the spectrum of $A \in \boldsymbol{C}^*$, as observed above). Thus, in the language of [10]:

3.7. Proposition. *The algebra C of complex numbers is a $\boldsymbol{K}$-codense cogenerator of the $\boldsymbol{K}$-category $\boldsymbol{C}^*$ of commutative C^*-algebras with identity.* Q.E.D.

In fact, for the Gelfand $\boldsymbol{K}$-monad $\boldsymbol{T}'$ in $\boldsymbol{A}'$ we also have $T'A \approx A$, via $\eta' A$, for all commutative C^*-algebras A, but it is clear that $\boldsymbol{T}'$ is not isomorphic to the identity on all of $\boldsymbol{A}'$.

Consider now the $\boldsymbol{K}$-category of $\boldsymbol{T}'$-algebras, that is, objects $A \in \boldsymbol{A}'$ provided with a $\boldsymbol{T}'$-algebra structure $T'A \xrightarrow{\alpha} A$, where $\alpha \circ \eta' A = \mathrm{id}$ and $T\alpha \circ \alpha = \mu' A \circ \alpha$, and maps: morphisms $A \xrightarrow{f} B$ in $\boldsymbol{A}'$ making the diagram

$$\begin{tikzcd} T'A \arrow[r, dashed, "\alpha"] \arrow[d, dashed, "T'\varphi"'] & A \arrow[d, dashed, "\varphi"] \\ T'B \arrow[r, "\beta"] & B \end{tikzcd}$$

commutative. Prop. 1.2 shows how $\boldsymbol{T}'$-algebras form a $\boldsymbol{K}$-category. (For a historical account and further references concerning these concepts see [12]; the enriched version used here is considered, for instance and among other places, in [6] and [10].) Denote this $\boldsymbol{K}$-category by $\boldsymbol{A}'^{T'}$. For any Kelley space $X\in\boldsymbol{K}$, $\bar{A}'(X,C)$ is a $\boldsymbol{T}'$-algebra with

$$T'\bar{A}'(X,C) = A'(A'(\bar{A}'(X,C),C),C) \xrightarrow{\bar{A}'(\epsilon X,C)=\alpha} \bar{A}'(X,C).$$

This is a general fact [10]. Given any C^*-algebra with identity A, the inverse of $\eta' A$ is a structure of T'-algebra on A, which is *trivial*: We will say that a $\boldsymbol{T}'$-algebra $T'A \xrightarrow{\alpha} A$ is trivial if α is a (two sided) inverse of $\eta' A$. Very naturally there arises the conjecture that all $\boldsymbol{T}'$-algebras are trivial. This is equivalent to saying that T' is idempotent $T'T' = \iota'$ (i.e. μ' is an isomorphism). It is also equivalent to the fact that the $\boldsymbol{K}$-functor $\boldsymbol{A}'^{T'} \xrightarrow{U^{T'}} \boldsymbol{A}'$, $U^{T'}(T'A \xrightarrow{\alpha} A) = A$ be $\boldsymbol{K}$-full-and-faithful, and, finally, equivalent to the (apparently weaker) fact that for any $\boldsymbol{T}'$-algebra, $T'A \xrightarrow{\alpha} A$, all maps into the complex numbers $A \to C$ in $\boldsymbol{A}'$ be morphism of $\boldsymbol{T}'$-algebras (see [10] Prop. II.4.6, p. 103). In fact, the conjecture would be true (in general) provided that is true for all $\boldsymbol{T}'$-algebras of the form $\bar{A}'(X,C)$ with $\alpha = \bar{A}'(\epsilon X, C)$, $X\in\boldsymbol{K}$. Recall that this is the case if X is completely regular and σ-compact (see Example 3.4 above). If there is an affirmative answer, a simple categorical $\boldsymbol{K}$-duality could be obtained as we shall see in Prop. 3.9. All the concrete $\boldsymbol{T}'$-algebras we have found are indeed trivial, but we have not succeeded, however, in proving this conjecture in general. Due to this unpleasant situation, we are forced to resort to a considerably more sophisticated machinery, developed in [1], in order to go on. The enriched version needed here, is actually to be found in [10].

The $\boldsymbol{K}$-duality produced below will give, as a byproduct the result that every K-topological algebra A satisfying the equivalent conditions of Prop. 3.1 is of the form $A = \bar{A}(X,C) = \bar{A}'(X,C)$ for some $X\in\boldsymbol{K}$. If our conjecture above is true, then necessarily $X = \boldsymbol{A}'(A,C)$. Before going to the general case, we will describe some sufficient conditions for this to be so.

3.8. Proposition. *If $A\in\boldsymbol{A}'$ satisfies the equivalent conditions in Prop. 3.1, then A has a canonical structure of $\boldsymbol{T}'$-algebra.*

Proof. We know that $A = \varprojlim_p A_p$ and that for each p, $A_p\in\boldsymbol{C}^*$. Therefore each A_p is a trivial $\boldsymbol{T}'$-algebra. Hence any morphism $A_p \to A_q$ in $\boldsymbol{A}'$ is a morphism of $\boldsymbol{T}'$-algebras; we can therefore take the limit $\varprojlim_p A_p$ *in* $\boldsymbol{A}'^{T'}$ (cf. [10] Propositions II.4.5 and

II.4.8). The T'-algebra so obtained can also be described by

$$
(1) \qquad
\begin{tikzcd}
T'A \arrow[rr, dashed, "\exists !\, \alpha"] \arrow[d, dashed, "T'\pi_p"'] \arrow[drr, phantom, "(1)"] & & A = \varprojlim_{p} A_p \arrow[d, "\pi_p"] \\
T'A_p \arrow[rr, "\approx", "(\eta' A_p)^{-1}"'] & & A_p
\end{tikzcd}
$$

Q.E.D.

3.9. Proposition. *Let $A \in \boldsymbol{A}'$ satisfy the equivalent conditions in Prop. 3.1 and assume furthermore that all morphisms $A \xrightarrow{\varphi} C$ in $\boldsymbol{A}$ are continuous in the locally m-convex topology defined by $\{p\}$. Then, the canonical $\boldsymbol{T}'$-algebra structure of A (Prop. 3.8) is trivial.*

Proof. Since $A \xrightarrow{\varphi} C$ is $\{p\}$-continuous we can assume, after replacing the family of seminorms by an equivalent family, if necessary, that φ is one of the projections $A \xrightarrow{\pi_p} A_p$ with $A_p = C$. Then, by definition (see diagram (1) above), φ is a morphism of $\boldsymbol{T}'$-algebras. Thus, any morphism $A \xrightarrow{\varphi} C$ in $\boldsymbol{A}$ is a morphism of $\boldsymbol{T}'$-algebras, and Prop. II.4.6 of [10] applies to complete the proof. Q.E.D.

3.10. Corollary. $A \approx T'A$ $(= \bar{\boldsymbol{A}}'(\boldsymbol{A}'(A,C),C))$ *in* $\boldsymbol{A}$. Q.E.D.

This corollary applies notably to the case of Fréchet algebras (see [27]), that is, to algebras satisfying (b) in Prop. 3.1 with a *countable* family $\{p\}$ and to arbitrary products of such algebras (cf. [31] Th. 5.2).

It is clear that K-topological algebras of the form $\bar{\boldsymbol{A}}'(X,C)$ for a general $X \in \boldsymbol{K}$ will not satisfy the assumptions in Prop. 3.9 (see 3.2.1 in Example 3.2), and this calls for a different approach. But before we describe it, let us observe the following.

3.11. Remark. *If $A \in \boldsymbol{A}$ has a structure of $\boldsymbol{T}$-algebra α, then A has an identity, i.e., $A \in \boldsymbol{A}'$. If both $A, B \in \boldsymbol{A}$ are $\boldsymbol{T}$-algebras, any morphism of $\boldsymbol{T}$-algebras $A \to B$ preserves the identity.*

Proof. The identity of A is $\alpha(e)$ where e is the identity of $TA = \boldsymbol{A}(\boldsymbol{A}(A,C),C)$; if $A \xrightarrow{\varphi} B$ in $\boldsymbol{A}$, $T(\varphi)$ always preserve the identity. Q.E.D.

Consider then the $\boldsymbol{K}$-functor $\boldsymbol{I} \xrightarrow{C} \boldsymbol{A}'$ and its codensity $\boldsymbol{K}$-monad $\boldsymbol{T}'$, that is, the Gelfand $\boldsymbol{K}$-monad. The following informal considerations are stated and proved in detail in [10], on pp. 135ff. under the heading "Second Relative V-Completion". The Gelfand $\boldsymbol{K}$-monad $\boldsymbol{T}'$ determines the $\boldsymbol{K}$-category $\boldsymbol{A}'^{\boldsymbol{T}'}$ of $\boldsymbol{T}'$-algebras, and $C \in \boldsymbol{A}'^{\boldsymbol{T}'}$, whence, we have a functor $\boldsymbol{I} \xrightarrow{C} \boldsymbol{A}'^{\boldsymbol{T}'}$. $\boldsymbol{A}'^{\boldsymbol{T}'}$ is a cotensored $\boldsymbol{K}$-category and then we have a codensity $\boldsymbol{K}$-monad in $\boldsymbol{A}'^{\boldsymbol{T}'}$. The cotensors in $\boldsymbol{A}'^{\boldsymbol{T}'}$ are (strictly) preserved by $\boldsymbol{A}'^{\boldsymbol{T}'} \xrightarrow{U^{\boldsymbol{T}}} \boldsymbol{A}'$, that is they are nothing but a $\boldsymbol{T}'$-algebra structure on $\bar{\boldsymbol{A}}'(X,C)$,

for $X \in \boldsymbol{K}$ (see the second paragraph after Prop. 3.7). The codensity $\boldsymbol{K}$-monad in $\boldsymbol{A}'^{T'}$ determines its own $\boldsymbol{K}$-category of algebras, which in turn, gives rise to a new $\boldsymbol{K}$-monad, and the whole process repeats itself again. After going up in this way an infinite number of times (once for each natural number), we have the (inverse) limit of the chain of $\boldsymbol{K}$-categories thus obtained. This category, due to the completeness of $\boldsymbol{K}$, is also a $\boldsymbol{K}$-category, which happens to be cotensored and "to contain" the object C (because C is coherently contained in each of the $\boldsymbol{K}$-categories in the chain). But then we have its codensity $\boldsymbol{K}$-monad, and this chain process starts again; and in this way we go through all the ordinals. *The limit of the (large!) chain just described (exists and) is also a $\boldsymbol{K}$-category*, call it $\boldsymbol{B}$, which is cotensored and such that C is a $\boldsymbol{K}$-codense cogenerator for $\boldsymbol{B}$: this means that the process stops. It is possible (and sometimes handy) to think on the objects of $\boldsymbol{B}$ as being those $\boldsymbol{K}$-topological algebras $A \in \boldsymbol{A}'$ which can be lifted all the way up, i.e., which admit a structure of algebra at every level in the chain. More accurately, in view of the possibility of different liftings, they should be considered as $\boldsymbol{K}$-topological algebras *together* with a structure of algebra at every level. *If a $\boldsymbol{K}$-topological algebra $A \in \boldsymbol{A}'$ is a trivial $\boldsymbol{T}'$ algebra (e.g., a commutative C^*-algebra) then it admits a unique lifting all the way up.* This is essentially due to the fact that the inverse of $\eta' A$ provides a (forced) lifting into $\boldsymbol{A}'^{T'}$ and that for the object so determined, the $\boldsymbol{K}$-monad in $\boldsymbol{A}'^{T'}$ is also trivial. This phenomenon is preserved in the steps corresponding to limit ordinals, too. We can summarize as follows:

3.12. Theorem. *There is a $\boldsymbol{K}$-complete $\boldsymbol{K}$-category $\boldsymbol{B}$ and a $\boldsymbol{K}$-faithful $\boldsymbol{K}$-functor $\boldsymbol{B} \xrightarrow{L} \boldsymbol{A}'$ which (strictly) preserves cotensors and $\boldsymbol{K}$-limits. Furthermore:*

(a) *there is a unique object $C \in \boldsymbol{B}$ such that $LC = C$; the diagram*

$$\begin{tikzcd} C \arrow[r] \arrow[d, "I"'] & \boldsymbol{B} \arrow[d, "L"] \\ C \arrow[r] & \boldsymbol{A}' \end{tikzcd}$$

commutes;

(b) *C is a $\boldsymbol{K}$-codense cogenerator of $\boldsymbol{B}$, that is, for all $B \in \boldsymbol{B}$, $B \approx \bar{\boldsymbol{B}}(\boldsymbol{B}(B, C), C)$;*

(c) *given any $A \in \boldsymbol{A}'$ such that $T'A \approx A$ via $\eta' A$, there is a unique object $B \in \boldsymbol{B}$ such that $LB = A$; and moreover, for any other $B' \in \boldsymbol{B}$, $\boldsymbol{B}(B, B') \approx \boldsymbol{A}'(A, LB')$ via L.*

The *proof* is to be found in [10]. Q.E.D.

We have the following corollary:

3.13. Theorem. *A $\boldsymbol{K}$-topological algebra $A \in \boldsymbol{A}'$ is of the form $\bar{\boldsymbol{A}}'(X, C)$ for some k-space X if and only if it satisfies the equivalent conditions in Prop. 3.1.*

Proof. We keep the notations of Prop. 3.1. The considerations made before Prop. 3.1

justify the "if" part. Assume now that $A = \varprojlim_p A_p$. From (c) in 3.12 there are (unique) $B_p \in \mathbf{B}$ such that $LB_p = A_p$, and we can take the limit $B = \varprojlim_p B_p$ in $\mathbf{B}$ (see Prop. 3.8). Since L preserves (strictly) limits, $LB = A$. On the other hand, it follows from (b) in 3.12 that $B \approx \bar{\mathbf{B}}(\mathbf{B}(B,C),C)$, and since L (strictly) preserves cotensors, we have $A = LB \approx L\bar{\mathbf{B}}(\mathbf{B}(B,C),C) = \bar{\mathbf{A}}'(\mathbf{B}(B,C),LC) = \bar{\mathbf{A}}'(\mathbf{B}(B,C),C)$. Thus, $A \approx \bar{\mathbf{A}}'(\mathbf{B}(B,C),C)$, and the proof is complete. Q.E.D.

Similar statements establishing functional representations for topological algebras can be found in [27] Theorem 8.4 and [19] Theorem 5. The result in the last corollary gives an isomorphism with an algebra of continuous complex functions in its natural Kelley topology, so that, in a sense, it can not be improved. However, if we adopt the customary standpoint of considering an algebra satisfying (a) and (b) in Prop. 3.1 qua locally m-convex algebra rather than as an element of $\mathbf{A}'$, the bijection $A \approx \bar{\mathbf{A}}'(X,C)$ is no longer a homeomorphism for the locally m-convex topologies on these algebras, but only continuous as $\bar{\mathbf{A}}'(X,C) \to A$: this accounts for the unpleasant asymmetry in the main result in [19].

It can easily be seen, following the proof of Prop. 3.9 that the functionals $A \xrightarrow{\varphi} C$ which are continuous for the locally m-convex topology $\{p\}$, can be lifted all the way up, or equivalently, they are morphisms at all levels, and therefore every such functional determines a point of $X = \mathbf{B}(B,C)$. This means that X contains the classical spectrum of $(A, \{p\})$, but might, a priori, be larger.

References

[1] H. Applegate and M. Tierney, Iterated cotriples, Lecture Notes No. 137 (Springer-Verlag, 1970).

[2] N. Bourbaki, Topologie générale, Chap. III and IV, A.S.I. 1143 (Hermann).

[3] N. Bourbaki, Intégration, Chap. 7 and 8, A.S.I. 1306 (Hermann, 1963).

[4] N. Bourbaki, Intégration, Chap. 1, 2, 3 and 4, A.S.I. 1175 (2nd ed., Hermann).

[5] N. Bourbaki, Théories spectrales, Chap. 1 and 2, A.S.I. (Hermann).

[6] M. Bunge, Relative functor categories and categories of algebras, J. Algebras 2 (1969) 61–101.

[7] B.J. Day and G.M. Kelley, Enriched functor categories, Lecture Notes No. 106 (Springer-Verlag, 1969).

[8] J. Dixmier, Les algèbres d'opérateurs dans l'espace Hilbertien (algèbres de von Neumann) (Gauthier-Villars, 1957).

[9] J. Dixmier, Les C^*-algèbres et leurs représentations (Gauthier-Villars, 1964).

[10] E.J. Dubuc, Kan extensions in enriched category theory, Lecture Notes No. 145 (Springer-Verlag, 1970).

[11] N. Dunford and J.T. Schwartz, Linear operators, I. (Interscience Publ., 1958).

[12] B. Eckmann (Ed.), Seminar on triples and categorical homology theory, Introduction, Lecture Notes No. 80 (Springer-Verlag, 1969).

[13] B. Eckmann and P. Hilton, Group-like structures in general categories I, Math. Ann. 145 (1962) 227–255.
[14] S. Eilenberg and G.M. Kelley, Closed categories, Proc. Conf. on Categorical algebra, La Jolla, 1965 (Springer-Verlag).
[15] P. Freyd, Abelian categories (Harper and Row, 1964).
[16] P. Gabriel and M. Zisman, Calculus of fractions and homotopy theory (Springer-Verlag, 1967).
[17] A. Guichardet et al., Special topics in topological algebras (Gordon and Breach, 1968).
[18] S. Hollard, A Hilbert space proof of the Banach-Steinhaus theorem, Am. Math. Monthly 76 (1969) 40–41.
[19] Do Shin Hsia, Seminormed Banach rings with involution, Izv. Akad. Nauk SSSR, Ser. Math. 23 (1959) 509–528 (Russian).
[20] S. Kakutani and V. Klee, The finite topology of a linear space, Arch. Math. 14 (1963) 55–58.
[21] J. Kelley, General topology (Van Nostrand, 1955).
[22] J. Kelley, I. Namioka et al., Linear topological spaces (Van Nostrand, 1963).
[23] G.M. Kelly Adjunction for enriched categories, Lecture Notes No. 106 (Springer-Verlag, 1969).
[24] F.E.J. Linton, Coequalizers in categories of algebras, Lecture Notes No. 80 (Springer-Verlag, 1969).
[25] F.E.J. Linton, On a pull-back lemma for Banach spaces and the functorial semantics of double negation, to appear.
[26] S. MacLane, Categorical algebra, Bull. Am. Math. Soc. 71 (1965) 40–106.
[27] E. Michael, Locally multiplicatively-convex topological algebras, Am. Math. Soc. Memoirs 11 (1952).
[28] M. Naimark, Normed rings (Nordhoff, 1964).
[29] J. Negrepontis, Applications of the theory of categories to analysis, Dissertation, McGill University, 1969.
[30] J. Negrepontis, Duality in analysis from the point of view of triples, to appear.
[31] N. Noble, k-spaces and some generalizations, Dissertation, University of Rochester, 1967.
[32] R. Pupier, Objets convenables. Objets de Kelley. Application aux espaces topologiques, Rev. Roum. Math. Pures et Appl. 14 (1969) 525–532.
[33] R. Pupier and A. Roux, Les espaces de Kelley et de DeGroot. Applications aux espaces localement compacts, Dept. Math., Fac. Sci. Univ. de Lyon, 1968.
[34] E. Spanier, Infinite symmetric products, function spaces, and duality, Ann. Math. 69 (1959) 142–198.
[35] N.E. Steenrod, A convenient category of topological spaces, Mich. Math. J. 14 (1967) 133–152.
[36] J. von Neumann, On a certain topology for a ring of operators, Ann. Math. 37 (1936) 111–115.
[37] D.D. Weddington, On k-spaces, Proc. Am. Math. Soc. 22 (1969) 635–638.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 4 (1971) 317–346

EQUATIONAL COMPLETION, MODEL INDUCED TRIPLES AND PRO-OBJECTS

J.F. KENNISON [1]
Clark University, Worcester, Massachusetts, U.S.A.

and

Dion GILDENHUYS [2]
McGill University, Montreal, Quebec, Canada

Received 6 February 1971

Introduction and preliminaries

We consider the following problem: Given a functor $I : \mathcal{M} \to \mathcal{S}$ (where $\mathcal{S}$ = sets), find the equational completion of I. That is to find an equational category $\mathcal{S}^{\mathbf{T}}$ together with a factorization of I through the underlying set functor (as $\mathcal{M} \to \mathcal{S}^{\mathbf{T}} \to \mathcal{S}$), which is "best possible" (so that if $\mathcal{M} \to \mathcal{S}^{\mathbf{T}_0} \to \mathcal{S}$ is another such factorization of I there exists a unique $U_0 : \mathcal{S}^{\mathbf{T}} \to \mathcal{S}^{\mathbf{T}_0}$ which makes everything commute). The existence (and a precise definition) of the equational completion if given by Linton [15] who shows the class of n-ary operations for the completion of I can be regarded as the class of all natural transformations from I^n to I where $n \in \mathcal{S}$. Thus the problem becomes, given $I : \mathcal{M} \to \mathcal{S}$, find some way of obtaining enough information about the natural transformations from I^n to I so that a reasonable description of $\mathcal{S}^{\mathbf{T}}$ can be given. A straightforward approach using the definition of a natural transformation is generally difficult because there often are many horribly infinitary operations for the equational completion. For example, if $\mathcal{M}$ = finite sets and $I : \mathcal{M} \to \mathcal{S}$ is the inclusion functor then a natural transformation from I^n to I corresponds to an ultrafilter on n (and the equational completion is the category of compact Hausdorff spaces). If $\mathcal{M}$ = fields the completion is the category of products of fields and continuous ring homomorphisms but the operations are difficult to des-

[1] J.F. Kennison thanks McGill University and the Canadian N.R.C. (through J. Lambek's grant) for support and encouragement. More recently Mr. Kennison has been supported by the N.S.F. under grant GP-14962.

[2] Dion Gildenhuys thanks the Canadian N.R.C. for support.

cribe usefully. (We do not know how to describe the category of algebras that results if one considers only the finitary operations.) Another difficulty with, say, simply making a list of natural transformations from I^n to I for various n is that it is hard to tell when one has listed enough operations to determine $\mathcal{S}^{\mathbf{T}}$. One can frequently find enough operations to determine a subtheory $\mathbf{T}_0$ of $\mathbf{T}$ such that every $M \in \mathcal{M}$ has the structure of a $\mathbf{T}_0$-algebra, the $\mathbf{T}_0$-homomorphisms are precisely the admissible homomorphisms (in $\mathcal{S}^{\mathbf{T}}$) between members of $\mathcal{M}$, and such that the $\mathbf{T}_0$-subalgebras and finite products will exactly determine the behaviour of subobjects and finite products in $\mathcal{S}^{\mathbf{T}}$ of (lifted) members of $\mathcal{M}$. However $\mathbf{T}_0$ might still be much smaller than $\mathbf{T}$. (For example if $\mathcal{M}$ = finite groups, I the obvious underlying functor, then $\mathbf{T}_0$ = the theory of groups has the above properties, but the completion $\mathbf{T}$ is the theory of profinite groups.)

Nonetheless once such a subtheory $\mathbf{T}_0$ has been found (cf. the definitions of separating triple and of normal separating triple and the procedure in Example 4.1) then one can often make effective use of a topological approach to $\mathcal{S}^{\mathbf{T}}$ based on triples and embed $\mathcal{S}^{\mathbf{T}}$ in the topological $\mathbf{T}_0$-algebras.

Fairly complete descriptions (Theorems 3.1, 3.2 and 3.3) of $\mathcal{S}^{\mathbf{T}}$ as a subcategory of topological $\mathbf{T}_0$-algebras are given if each $M \in \mathcal{M}$ is finite or satisfies a descending chain type of condition. In these cases $\mathcal{S}^{\mathbf{T}}$ is related to the Pro-objects for $\mathcal{M}$ (which are briefly reviewed in Section 2). Examples involving pseudocompact rings, modules and algebras and the case $\mathcal{M}$ = countable sets show that the topological approach works well in specific cases too.

The equational completions that we shall examine turn out to be varietal, hence arise from triples on $\mathcal{S}$. If $\mathcal{S}^{\mathbf{T}}$ is the completion of $I: \mathcal{M} \to \mathcal{S}$ we shall regard $\mathbf{T}$ as a triple (rather than as a varietal equational theory) in which case $\mathbf{T}$ is the model induced triple arising from $I: \mathcal{M} \to \mathcal{S}$. (This follows from the argument sketched below and from the results of [1] where the dual notion, of a model induced cotriple is defined. These triples are called codensity triples in [14] and their relationship to equational completions is there established in a general setting. Therefore this paper gives some techniques for computing examples of the categories of algebras discovered by Linton, Appelgate and Tierney when we consider relatively specific cases. Incidentally we shall also relate $\mathcal{S}^{\mathbf{T}}$ to a model induced triple over Top (= Topological spaces and maps).)

Let us recall that if $\mathbf{T} = (T, \eta, \mu)$ is the model induced triple for $I: \mathcal{M} \to \mathcal{S}$ then $T(n) = \lim [(n, I) \to \mathcal{S}]$ where $n \in \mathcal{S}$ and (n, I) is the comma category of all functions $n \to I(M)$ where M ranges over the class of models (i.e. objects of $\mathcal{M}$). The functor $(n, I) \to \mathcal{S}$ assigns $I(M)$ to the function $n \to I(M)$. (This limit exists iff $\mathbf{T}$ is well-defined iff I is tractable (i.e. the collection n.t. (I^n, I) is small) iff the equational completion is varietal. A practical test for $I: \mathcal{M} \to \mathcal{S}$ to be tractable is given in the discussion preceeding 1.3 below.) For each $g: n \to I(M)$ in (n, I) we let $\langle g \rangle: T(n) \Rightarrow I(M)$ be the corresponding projection. Then η and μ are defined by $\langle g \rangle \eta = g$ and $\langle g \rangle \mu = \langle\langle g \rangle\rangle$. If $f: n \to m$ then $T(f)$ is defined by $\langle g \rangle T(f) = \langle gf \rangle$. Moreover every natural transformation $\lambda: I^n \to I$ gives rise to $\lambda \in T(n)$ where $\langle g \rangle(\lambda) = \lambda_M(g)$ for

$g: n \to I(M)$, that is for $g \in I^n(M)$. Thus every such natural transformation λ gives rise to an n-ary operation λ of T. This correspondence in one-one and onto and illustrates why the model induced triple gives rise to the equational completion. (We shall generally speak of n-ary operations of triples when strictly speaking we mean an n-ary operation of the corresponding varietal theory.)

In general if $\mathbf{T}$ is a triple on $\mathcal{A}$ then the statement $(A, \theta) \in \mathcal{A}^{\mathbf{T}}$ or (A, θ) is a $\mathbf{T}$-algebras means that $A \in \mathcal{A}$ and $\theta : T(A) \to A$ is a structure map. The morphisms of $\mathcal{A}^{\mathbf{T}}$ shall be referred to as $\mathbf{T}$-homomorphisms or as morphisms of $\mathcal{A}$ which are admissible. If $\mathbf{T}_0$ is a triple over sets then a topological $\mathbf{T}_0$-algebra with a topology such that the n-ary operations are continuous, using the product topology.

We shall use the term "limit" to refer to generalized inverse limits (i.e. the left roots in [6]). For emphasis, limits in $\mathcal{S}^{\mathbf{T}}$ shall sometimes be called $\mathbf{T}$-limits. Limits in Top shall sometimes be referred to as Top-limits. We say that a small category D is *filtered* if $d, e \in D$ imply there exists $c \in D$ and morphisms from c to d and from c to e. Also if $f, g \in D(c, d)$ then there exists h with $fh = gh$. A *filtered diagram* is a functor whose domain is filtered and its limit is a *filtered limit*. (Aside from the use of contravariant functors in [2], this definition is effectively equivalent to the definition of filtered limit in [2].) If $\mathcal{A}$ is a category then $\mathcal{A}^{\mathrm{op}}$ is the dual category and $\mathcal{A}(X, Y)$ is the set of morphisms from X to Y. If X is an object of $\mathcal{A}$ then X also denotes the identity morphism in $\mathcal{A}(X, X)$. If $\mathcal{A}$ and $\mathcal{B}$ are categories then $(\mathcal{A}, \mathcal{B})$ denotes the possibly illegitimate category of functors from $\mathcal{A}$ to $\mathcal{B}$. A subcategory $\mathcal{A}$ of $\mathcal{B}$ is *reflective* if the inclusion functor $\mathcal{A} \to \mathcal{B}$ has a left adjoint. (This follows Freyd [6].) A morphism is a *split epi* if it has a right inverse. The term *quotient map* is always used in the *topological* sense. Parentheses in expressions such as $T(n)$ are sometimes omitted when they are not needed particularly in complicated formulas (when only the crucial parentheses are included).

The authors would like to express their thanks to Michael Barr for a stimulating seminar on triples given at McGill University. Among other things he posed the problem of finding the equational completion of the category of fields. We also wish to thank Myles Tierney for a helpful conversation concerning pro-objects and $\mathcal{M}$-objects.

§ 1. Some topological observations

Notation. From here on $I : \mathcal{M} \to \mathcal{S}$ shall denote a tractable functor and $\mathbf{T} = (T, \eta, \mu)$ shall be the model induced triple on $\mathcal{S}$. The category of $\mathbf{T}$-algebras shall be $\mathcal{S}^{\mathbf{T}}$ with $U_{\mathbf{T}} : \mathcal{S}^{\mathbf{T}} \to \mathcal{S}$ and $F_{\mathbf{T}} : \mathcal{S} \to \mathcal{S}^{\mathbf{T}}$ the underlying and free functors.

Given $M \in \mathcal{M}$ then $I(M) \in \mathcal{S}^{\mathbf{T}}$ shall be defined by the structure map $\langle TM \rangle : TIM \to IM$ (here IM denotes the identity map on IM — this notation for identity map is used throughout). This defines the lifted model functor $I : \mathcal{M} \to \mathcal{S}^{\mathbf{T}}$.

If there is no danger of confusion and if $M \in \mathcal{M}$ we shall also use M to denote $I(M) \in \mathcal{S}$ and $I(M) \in \mathcal{S}^{\mathbf{T}}$.

The topology on a $\mathbf{T}$*-algebra.* Let $M \in \mathfrak{M}$. Then $T(n)$ can be topologized by the smallest topology rendering each $\langle g\rangle : T(n) \to M$ continuous, when each M is given the discrete topology. (Thus the $\mathcal{S}$-limit, $T(n)$, is regarded as a topological limit of discrete spaces.)

If $(n, \theta) \in \mathcal{S}^{\mathbf{T}}$ (so that $\theta : T(n) \to n$ is the structure map) we define $Q(n, \theta)$ as the quotient topology induced by θ. Then $Q : \mathcal{S}^{\mathbf{T}} \to$ Top is easily seen to be a functor. Let U : Top $\to \mathcal{S}$ be the underlying functor and $D : \mathcal{S} \to$ Top the left adjoint, which assigns the discrete topology to a set. Let $I_D : \mathfrak{M} \to$ Top be defined by $I_D = DI (= QI)$ and let $\widetilde{\mathbf{T}} = (\widetilde{T}, \widetilde{\eta}, \widetilde{\mu})$ be the model induced triple on Top. Let $\widetilde{U}$: Top$^{\widetilde{\mathbf{T}}} \to$ Top, $\widetilde{F}$: Top $\to$ Top$^{\widetilde{\mathbf{T}}}$, $\widetilde{I} : \mathfrak{M} \to$ Top$^{\widetilde{\mathbf{T}}}$ be the obvious functors. For convenience, if $X \in$ Top and $g : X \to M$ is continuous (that is $g : X \to I_D(M)$) we let $[g] : \widetilde{T}(X) \to M$ be the corresponding projection which makes $\widetilde{T}(X)$ the limit of $(X, I_D) \to$ Top. There exists a comparison functor φ : Top$^{\widetilde{\mathbf{T}}} \to \mathcal{S}^{\mathbf{T}}$ since the adjointness $\widetilde{F}D \dashv U\widetilde{U}$ generates the triple $\mathbf{T}$. Finally, if $M \in \mathfrak{M}$ we shall also use M to denote $I_D(M) \in$ Top and $\widetilde{I}(M) \in$ Top$^{\widetilde{\mathbf{T}}}$ (as well as $I(M) \in \mathcal{S}$ and $I(M) \in \mathcal{S}^{\mathbf{T}}$) so long as the context makes the exact meaning clear. To summarize the above and to record some immediate properties of Q we state:

1.1. Proposition. (*We suggest drawing a diagram of the above categories and functors. For clarity the category* $\mathfrak{M}$ *and the functors* $I, \widetilde{I}, \widetilde{I}_D$ *and* I_D *and* $\widetilde{I}$ *might be omitted.*)

(a) $U_{\mathbf{T}}\varphi = U\widetilde{U}$ *and* $\varphi\widetilde{F}D = F_{\mathbf{T}}$ *and* φ *preserves limits.*

(b) Q *preserves quotients (meaning that if* $f : A \to B$ *is an onto* $\mathbf{T}$*-homomorphism then* $Q(f)$ *is a quotient map in the topological sense).*

Q *preserves the topology of* $T(n)$ *(meaning that* $Q(T_n, \mu)$ *is the limit topology on* Tn *mentioned above).*

However, Q *need not assign the relative topology to a* $\mathbf{T}$*-subalgebra nor does* Q *generally preserve limits nor is* $Q\varphi$ *necessarily equal to* $\widetilde{U}$. *We can say that* $Q(\lim A_i)$ *is at least as large (has at least as many open sets) as* Top-lim $Q(A_i)$, *that* $Q\varphi(\widetilde{X})$ *is at least as large in* $\widetilde{U}(\widetilde{X})$ *for* $\widetilde{X} \in$ Top$^{\widetilde{\mathbf{T}}}$ *and that* $Q(A)$ *is at least as large as the relative topology on* A *induced by* $Q(B)$ *if* $A \subseteq B$ *is a* $\mathbf{T}$*-subalgebra.*

(c) *If* $X \in$ Top *then* $\exists\lambda(=\lambda_X) : TUX \to \widetilde{T}X$ *defined by* $[g]\,\lambda = \langle Ug\rangle$. *Then* λ *is continuous and can be regarded as a natural transformation from* $QF_{\mathbf{T}}U \to \widetilde{T}$. *If* $(X, \theta) \in$ Top$^{\widetilde{\mathbf{T}}}$ *then* $\varphi(X, \theta) = (UX, \theta\lambda_X)$. *Moreover if* X *is discrete then* λ_X *is the identity.*

Proof. (a) follows from the construction of φ, see [4].

As for (b) we first note that $\mu : T^2(n) \to T(n)$ is a quotient map as it is split epi in Top with $T(\eta)$ as right inverse. Thus $Q(Tn, \mu)$ has the limit topology. Now if $f : (X, \theta) \to (Y, \psi)$ is onto and admissible then $f\theta = \psi\, Tf$ and Tf is a quotient map (in fact Tf is split epi in Top as f is split epi in $\mathcal{S}$). Thus $f\theta$ is a quotient map and θ is continuous, so f is a quotient map.

Next $Q(\lim A_i)$ has at least as many open sets as Top-lim $Q(A_i)$ as the projections

are Q-continuous. Similarly the Q topology on A is at least as large as the relative topology as the inclusion $A \to \underline{B}$ is Q-continuous.

Finally let $\widetilde{X} = (X, \theta) \in \mathrm{Top}^{\mathbf{T}}$. By (c), $\varphi(X, \theta) = (UX, \theta\lambda_X)$. Clearly $\theta\lambda_X$ is continuous from TUX to $\widetilde{UX}$. Hence $Q\varphi(X, \theta)$, the space with the quotient topology induced by $\theta\lambda_X$ has at least as many open sets as $\widetilde{UX}$.

(c) Since $\eta : UX \to TUX$ is a front adjunction there is a unique $\lambda : TUX \to \widetilde{T}X$ which is $\mathbf{T}$-admissible and satisfies $\lambda\eta = \widetilde{\eta}$. Note that $|g|\lambda = \langle Ug \rangle$ follows. Also $\theta\lambda$ is the structure map of (X, θ) as $\theta\lambda$ is $\mathbf{T}$-admissible and $\theta\lambda\eta = X$.

1.2. Proposition. *(Continuity of the operations.) Let $\mathbf{T}_0$ be any equational theory generated by some finitary operations of $\mathbf{T}$. Then:*

(a) *If $(X, \theta) \in \mathcal{S}^{\mathbf{T}}$ and if $\theta^k : |T(X)|^k \to X^k$ is a quotient map (between the product topologies) for all finite k, then (X, θ), with the Q topology, is a topological $\mathbf{T}_0$-algebra. (In general, products of quotient maps need not be quotient maps. However if $\mathbf{T}$ admits a group operation or if every model is finite then θ^k will be a quotient map.)*

(b) *If ω is any k-ary operation of $\mathbf{T}$, for k finite, and if $A \in \mathcal{S}^{\mathbf{T}}$ then ω is continuous from $Q(A^k) \to Q(A)$ (where A^k is the product algebra in $\mathcal{S}^{\mathbf{T}}$. Note that $Q(A^k)$ may fail to be $Q(A)^k$.)*

(c) *Each $\widetilde{\mathbf{T}}$-algebra is a topological $\mathbf{T}_0$-algebra using the $\widetilde{U}$-topology, and is Hausdorff.*

Remark. The infinitary operations of $\mathbf{T}$ are usually not continuous.

Proof. (a) $T(X)$ is always a topological $\mathbf{T}_0$-algebra as it is a topological and algebraic limit of models which are (discrete) topological $\mathbf{T}_0$-algebras. Thus the operations of $\mathbf{T}_0$ are continuous from $|T(X)|^k$ to $T(X)$. If θ^k is a quotient map then clearly all $\mathbf{T}_0$-operations are continuous from $X^k \to X$ (by the naturality of operations).

If $\mathbf{T}$ admits a group operation then a well known argument shows that θ is an open mapping. (Usually the argument is stated in the presence of Hausdorffness which however is not needed.) If θ is an open onto mapping then θ^k is a quotient map for all k.

If every model if finite then the Q topology is always compact, Hausdorff (as shown in the proof of 3.1) and every continuous onto map is closed hence a quotient map.

(b) Notice that there exists a continuous map from $T(A^k) \to |T(A)|^k$ whose projections are $T(A^k \to A)$. But every k-ary operation gives rise to a continuous map from $T(A)^k \to T(A)$ [as in (a)] hence there exists a continuous map from $T(A^k) \to T(A)$. By taking quotients one can readily show that the operation is continuous from $Q(A^k)$ to $Q(A)$.

(c) Let $(X, \widetilde{\theta})$ be a $\widetilde{\mathbf{T}}$-algebra, where $\widetilde{\theta} : \widetilde{T}(X) \to X$. Then $\widetilde{\theta}$ induces a $\mathbf{T}_0$-structure on X (which coincides with the underlying $\mathbf{T}_0$-structure of $\varphi(X, \theta)$, see a similar discussion in 1.4). Moreover $\widetilde{U}(\widetilde{\theta})$ is split epi in Top as $\widetilde{\eta}$ is continuous. Hence $\widetilde{\theta}$ and $\widetilde{\theta}^k$ are quotient maps for all k and the above arguments apply.

Since $\tilde{\theta}$ is split epi, $\tilde{U}(X)$ is topologically equivalent to a subspace of $\tilde{T}(X)$ hence is Hausdorff.

Definition. $\mathbf{T}_0 = (T_0, \eta_0, \mu_0)$ is a *separating triple* for $I : \mathcal{M} \to \mathcal{S}$ if each model has the structure of a $\mathbf{T}_0$-algebra; the models are closed under the formation of $\mathbf{T}_0$-sub-algebras and the maps between models are precisely the $\mathbf{T}_0$-homomorphisms. In more precise terms $\mathbf{T}_0$ is a separating triple for I if I factors as $\mathcal{M} \to \mathcal{S}^{\mathbf{T}_0} \to \mathcal{S}$ where $\mathcal{M} \to \mathcal{S}^{\mathbf{T}_0}$ embeds $\mathcal{M}$ as a full subcategory closed under the formation of $\mathbf{T}_0$-sub-algebras, and where $\mathcal{S}^{\mathbf{T}_0} \to \mathcal{S}$ is the underlying set functor.

We say that $\mathbf{T}_0$ is *finitary* if the corresponding equational theory is (i.e. if it is generated by the finitary operations).

Notation and remarks. If $\mathbf{T}_0$ is a separating triple and $n \in \mathcal{S}$ then we let $(n, I)_0$ be the full subcategory of (n, I) consisting of those $f : n \to M$ such that $f(n)$ generates M as a $\mathbf{T}_0$-algebra (i.e. the extension $T_0(n) \to M$ is onto). We observe:
(1) If $f_1 : n \to M_1$ and $f_2 : n \to M_2$ are in $(n, I)_0$ and if $e : f_1 \to f_2$ (that is $e : M_1 \to M_2$ and $ef_1 = f_2$), then e is onto as its range contains $f_2(n)$ which generates M_2. Also e is unique as it is determined on $f_1(n)$. Thus $(n, I)_0$ is partially ordered.
(2) $(n, I)_0$ is initial in (n, I) and $T(n) = \lim\ [(n, I)_0 \to \mathcal{S}]$. Moreover, $(n, I)_0$ is always small so $T(n)$ exists hence $I : \mathcal{M} \to \mathcal{S}$ is automatically tractable if a separating triple exists.
(3) Suppose that $\mathcal{M}$ has and I preserves finite products. Let $f_1 : n \to M_1$ and $f_2 : n \to M_2$ be in (n, I). Let $(f_1, f_2) : n \to M_1 \times M_2$ be the obvious map. Let M be the $\mathbf{T}_0$-subalgebra generated by the range of (f_1, f_2). Denote by $f_1 \wedge f_2 : n \to M$ the map induced by (f_1, f_2). Then $f_1 \wedge f_2 \in (n, I)_0$ and is the inf of f_1 and f_2 whenever $f_1, f_2 \in (n, I)_0$.
(4) If $\zeta \in T(n)$ then the open sets of the form $\langle g \rangle^{-1}(m)$ for $g : n \to M$ in $(n, I)_0$ and $m = \langle g \rangle(\zeta)$ form a base for the neighborhoods at ζ. (Since $T(n)$ has the limit topology these neighborhoods form a subbase at ζ and they are also closed under finite intersections in view of the construction in (3) above.) This argument clearly applies to any filtered topological limit of discrete spaces.

1.3. Proposition. *(Consequences of a separating triple.) Let $\mathbf{T}_0$ be a separating triple for $I : \mathcal{M} \to \mathcal{S}$ (thus I is tracable by (2) above). Assume that $\mathcal{M}$ has and I preserves finite products. Then:*

(a) *There is a natural map $t : T_0(X) \to T(X)$ which has dense range for all $X \in \mathcal{S}$. If $(X, \theta) \in \mathcal{S}^{\mathbf{T}}$ then $(X, \theta t) \in \mathcal{S}^{\mathbf{T}_0}$ and is the underlying $\mathbf{T}_0$-algebra of (X, θ).*

(b) *If (X, θ) and (Y, ψ) are $\mathbf{T}$-algebras and $Q(Y, \psi)$ is Hausdorff then a $\mathbf{T}$-homomorphism from (X, θ) to (Y, ψ) is the same thing as a continuous $\mathbf{T}_0$-homomorphism.*

(c) *A closed $\mathbf{T}_0$-subalgebra of a $\mathbf{T}$-algebra is a $\mathbf{T}$-subalgebra.*

(d) $\mathrm{Top}^{\mathbf{T}}$ *is co-complete (as well as complete) so $\varphi : \mathrm{Top}^{\mathbf{T}} \to \mathcal{S}^{\mathbf{T}}$ has a left adjoint. Moreover if $Q\varphi = \tilde{U}$ then φ is a full embedding.*

(e) *If* $\mathbf{T}_0$ *is finitary then* $\mathrm{Top}^{\tilde{\mathbf{T}}}$ *can be fully embedded into the topological* $\mathbf{T}_0$*-algebras (using the* $\tilde{U}$ *topology).*

Proof. Since $T(X)$ is a filtered limit of models over $(X, I)_0$ (which is filtered by (3) above) a basic neighbourhood of $\zeta \in T(X)$ has the form $\langle g \rangle^{-1}(m)$ where $m = \langle g \rangle(\zeta)$ and $g \in (X, I)_0$. Given $g : X \to M$ in $(X, I)_0$ let $g : T_0(X) \to M$ be the extension (so $g\,\eta_0 = g$). Define t by $\langle g \rangle t = g$ for all $g \in (X, I)_0$. Since g is always onto, the range of t meets all basic neighbourhoods of each $\zeta \in TX$. Thus t has dense range. It is clear that $t : T_0 \to T$ can be regarded as a natural transformation. Finally we know that the underlying $\mathbf{T}_0$-algebra of TX is $\mathbf{T}_0$-lim $[(X, I)_0 \to \mathscr{S}^{\mathbf{T}_0}]$ as limits are preserved. Thus $t : T_0(X) \to T(X)$ and θ are (underlying) $\mathbf{T}_0$-homomorphisms. Moreover $t\,\eta_0 = \eta$ so $\theta\, t\, \eta_0 = X$ and θt is therefore the structure map of the $\mathbf{T}_0$-algebra underlying (X, θ).

As for (b) if $f : (X, \theta) \to (Y, \psi)$ is a continuous $\mathbf{T}_0$-homomorphism then $f\theta t = \psi\, T(f)t$ so $f\theta = \psi\, T(f)$ as t has dense range and $Q(Y, \psi)$ is Hausdorff.

As for (c) let (X, θ) be in $\mathscr{S}^{\mathbf{T}}$ and let $A \subset X$ be a closed $\mathbf{T}_0$-subalgebra. Let $i : A \to X$ be the inclusion and $\psi_0 : T_0(A) \to A$ the $\mathbf{T}_0$-structure map. It suffices to show that the range of $\theta\, T(i)$ is continued in A. But $\theta\, T(i) t_A = \theta t_X T_0(i) = i\psi_0$, hence $\theta T(i) t_A$ has range in A. But the range of $\theta T(i) t_A$ is dense in the range of $\theta T(i)$ (as t_A has dense range) and A is closed so that range of $\theta T(i)$ is contained in A.

As for (d) we observe that if $(X, \theta) \in \mathrm{Top}^{\tilde{\mathbf{T}}}$ then $\tilde{U}(X, \theta) = X$ is Hausdorff as it is a retract (via $\tilde{\eta}$ and θ) of $\tilde{T}(X)$. Also $\lambda : TU(X) \to \tilde{T}(X)$ has dense range for the same reason that t does. Thus by the above arguments a $\tilde{U}$-closed $\mathbf{T}_0$-subalgebra of a $\mathbf{T}$-algebra is a $\tilde{\mathbf{T}}$-subalgebra. Thus $\mathrm{Top}^{\tilde{\mathbf{T}}}$ has coequalizers for if $g, h : Y \to Z$ are $\tilde{\mathbf{T}}$-homomorphisms then the set of all e with dense range and $eg = eh$ is a solution set. Thus $\mathrm{Top}^{\tilde{\mathbf{T}}}$ is cocomplete (by [14]) and φ has a left adjoint (by [4]).

Also the above arguments show that a $\tilde{\mathbf{T}}$-homomorphism is the same thing as a $\tilde{U}$-continuous $\mathbf{T}_0$-homomorphism. So φ is obviously full if $Q\varphi = \tilde{U}$. By construction φ is faithful.

As for (e) every $\tilde{\mathbf{T}}$-algebra can be regarded as a topological $\mathbf{T}_0$-algebra by 1.2(c) The anal'g of (b) shows that this is a full embedding. q.e.d.

We are particularly interested in the case when $\mathbf{T}_0$ admits a group operation (which will usually be denoted multiplicatively). In this case K is said to be a $\mathbf{T}_0$-*kernel* of $X \in \mathscr{S}^{\mathbf{T}_0}$ if there exists a $\mathbf{T}_0$-homomorphism f with $K = f^{-1}(1)$ (where 1 denotes the group identity). As usual a $\mathbf{T}_0$-quotient of X is an (onto) image of X in $\mathscr{S}^{\mathbf{T}_0}$. Thus a $\mathbf{T}_0$-quotient of X can be represented as X/K for K a $\mathbf{T}_0$-kernel. The following lemma is useful.

1.4. Lemma. *Let* $\mathbf{T}_0$ *be a triple over* $\mathscr{S}$ *which admits a group operation. Then the notions of* $\mathbf{T}_0$*-kernel and* $\mathbf{T}_0$*-quotient are defined above and satisfy the following properties:*

(a) *Let X and Y be $\mathbf{T}_0$-algebras and $f: X \to Y$ a $\mathbf{T}_0$-homomorphism. If $K \subseteq Y$ is a $\mathbf{T}_0$-kernel then so is $f^{-1}(K)$. If f is onto and if $N \subseteq X$ is a $\mathbf{T}_0$-kernel then so is $f(N)$.*

(b) *If K and N are $\mathbf{T}_0$-kernels of X then so is KN.*

(c) *Let X be a topological $\mathbf{T}_0$-algebra. Then KU is an open $\mathbf{T}_0$-kernel if U is an open $\mathbf{T}_0$-kernel and K is any $\mathbf{T}_0$-kernel. Moreover if K is a closed $\mathbf{T}_0$-kernel and if X admits a base $\{U_\alpha\}$ of open $\mathbf{T}_0$-kernel neighbourhoods of 1 then $K = \bigcap KU_\alpha$.*

Proof. Let N be a normal subgroup of the $\mathbf{T}_0$-algebra X. Then N is a $\mathbf{T}_0$-kernel iff the equivalence relation, mod N, is a $\mathbf{T}_0$-congruence (i.e. is compatible with every operation of $\mathbf{T}_0$. We are regarding $\mathbf{T}_0$ as an equational theory). (a) is an immediate consequence of this observation, and the argument below.

As for (b) let w be an n-ary operation of T_0 and let $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ be n-tuples of X with $x_i \equiv y_i$ mod KN for all i. (Despite our notation we do not assume that n is finite.) Then there exist $k_i \in K$ and $n_i \in N$ with $x_i = k_i n_i y_i$ for all i. Let $z_i = n_i y_i$, then $w(x_i) \equiv w(z_i)$ as K is a $\mathbf{T}_0$-kernel and $w(z_i) \equiv w(y_i)$ as N is a $\mathbf{T}_0$-kernel.

(c) is obvious. q.e.d.

1.5. Proposition. *Let $\mathbf{T}_0$ be a separating triple with a group operation. Assume that $\mathcal{M}$ has and I preserves finite products. Then:*

(a) *The discrete members of $\mathcal{S}^{\mathbf{T}}$ are precisely the $\mathbf{T}_0$-quotients of models. (That is if M is any model and K any $\mathbf{T}_0$-kernels then $M/K \in \mathcal{S}^{\mathbf{T}}$ the projection $M \to M/K$ is a $\mathbf{T}$-homomorphism, $Q(M/K)$ is discrete and every discrete $\mathbf{T}$-algebra is of this form.)*

(b) *Let $(X, \theta) \in \mathcal{S}^{\mathbf{T}}$. (By 1.2 the Q topology makes (X, θ) a topological group.) The open $\mathbf{T}_0$-kernels form a basic system of neighbourhoods at the group identity $1 \in X$.*

Moreover $U \subseteq X$ is an open $\mathbf{T}_0$-kernel iff U is a closed $\mathbf{T}_0$-kernel and X/U is $\mathbf{T}_0$-equivalent to a $\mathbf{T}_0$-quotient of a model, which is true iff U is a $\mathbf{T}$-kernel and X/U is $\mathbf{T}$-equivalent to a $\mathbf{T}_0$-quotient of a model.

(c) *A closed $\mathbf{T}_0$-kernel of a $\mathbf{T}$-algebra is a $\mathbf{T}$-kernel.*

Remark. If we increase $\mathcal{M}$ by closing up under the formation of $\mathbf{T}_0$-quotients, this will not change $\mathcal{S}^{\mathbf{T}}$, in view of (a), and will not change the basic neighbourhoods of (b) so Q will not be affected either. Note that the hypotheses of this proposition are also preserved. Hence we shall often assume that $\mathcal{M}$ is closed under the formation of $\mathbf{T}_0$-quotients in cases when this entails no real loss of generality. Then the above statements are somewhat simplified.

Proof. (a) If M is a model and K is a $\mathbf{T}_0$-quotient of M then $M/K \in \mathcal{S}^{\mathbf{T}}$, the projection $M \to M/K$ is a $\mathbf{T}$-homomorphism and $Q(M/K)$ is discrete in view of the Lemma 1.6, below. Conversely, every discrete $\mathbf{T}$-algebra is of this form in view of (b) proven below.

(b) Let N be a neighbourhood of $1 \in X$. Then $\theta^{-1}(N)$ is a neighbourhood of $1 \in TX$ so there exists $h : X \to M$ with $\mathrm{Ker}\langle h\rangle \subseteq \theta^{-1}(N)$. Thus $\theta(\mathrm{Ker}\langle h\rangle) \subseteq N$ and $\theta(\mathrm{Ker}\langle h\rangle)$ is an open $\mathbf{T}_0$-kernel as θ is an onto, open mapping. Thus the open $\mathbf{T}_0$-kernels are a base at 1.

Next let $U \subseteq X$ be an open $\mathbf{T}_0$-kernel. Then $\theta^{-1}(U)$ is an open neighbourhood of $1 \in T(X)$. Using the fact that $T(X)$ is a filtered limit of models there exists $h : X \to M$ in $(X, I)_0$ with $\mathrm{Ker}\langle h\rangle \subseteq \theta^{-1}(U)$. Since $\langle h\rangle$ is onto, $\langle h\rangle\theta^{-1}(U)$ is a $\mathbf{T}_0$-kernel of M by 1.4(a). Let $K = \langle h\rangle\theta^{-1}(U)$. Let $p : M \to M/K$ be the projection, then p is a $\mathbf{T}$-homomorphism by (a). Note that $\mathrm{Ker}\, p\langle h\rangle = \theta^{-1}(U)$ (since $\mathrm{Ker}\, p\langle h\rangle = \langle h\rangle^{-1}\langle h\rangle(\theta^{-1}U)$ and $\mathrm{Ker}\langle h\rangle \subseteq \theta^{-1}(U)$). Thus $U = \theta(\mathrm{Ker}\, p\langle h\rangle)$ is a $\mathbf{T}$-kernel by applying 1.4(a) to $\mathbf{T}$. Hence X/U is equivalent to M/K as they are both quotients of TX by the same kernel.

Conversely let U be a closed $\mathbf{T}_0$-kernel with X/U $\mathbf{T}_0$-equivalent to a $\mathbf{T}_0$-quotient of a model. By (c) proved below, U is a $\mathbf{T}$-kernel so $X/U \in \mathcal{S}^{\mathbf{T}}$. Let $f : M/K \to X/U$ be the $\mathbf{T}_0$-equivalence where $M \in \mathcal{M}$. By (a), M/K is discrete hence f is continuous. Since U is closed and since $Q(X/U)$ has the quotient topology, $Q(X/U)$ is Hausdorff. Thus by 1.3(b), f is a $\mathbf{T}$-homomorphism. Since f is one-one and onto it is a $\mathbf{T}$-equivalence. Thus $Q(X/U) \approx Q(M/K)$ which is discrete and U is open. The other characterisation of open $\mathbf{T}_0$-kernels follows readily.

(c) Let K be a closed $\mathbf{T}_0$-kernel of X and let $\{U_\alpha\}$ be the set of open $\mathbf{T}_0$-kernels. Then $K = \bigcap KU_\alpha$ and each KU_α is an open $\mathbf{T}_0$-kernel (hence a $\mathbf{T}$-kernel) by 1.4(b) and (c). By considering products it follows that intersections of $\mathbf{T}$-kernels are $\mathbf{T}$-kernels, so K is a $\mathbf{T}$-kernel. q.e.d.

1.6. Lemma. *Let $\mathbf{T}_0$ be a separating triple and assume that $\mathcal{M}$ has and I preserves finite products. Let (X, θ) be a $\mathbf{T}$-algebra (then (X, θ) has a topology supplied by Q and an underlying $\mathbf{T}_0$-structure). Let (Y, ψ_0) be a $\mathbf{T}_0$-algebra and $f : X \to Y$ an onto $\mathbf{T}_0$-homomorphism. Assume that the quotient topology on Y induced by f is Hausdorff. Then there exists a unique structure map ψ with (Y, ψ) a $\mathbf{T}$-algebra and f a $\mathbf{T}$-homomorphism. It necessarily follows that (Y, ψ) is compatible with (Y, ψ_0) and that $Q(Y, \psi)$ coincides with the quotient topology induced by f.* (Note that in effect this lemma gives sufficient conditions for a $\mathbf{T}_0$-congruence to be a $\mathbf{T}$-congruence.)

Proof. Since f is a $\mathbf{T}_0$-homomorphism, $\psi_0 T_0(f) = f\theta\, \iota_X$. Let $e : Y \to X$ be any right inverse of f (so $fe = Y$) and define $\psi = f\theta\, T(e)$. Then $\psi\iota_Y = \psi_0 T_0(f) T_0(e) = \psi_0$. Moreover a density argument shows that $\psi T(f) = f\theta$. Using this fact and the facts that f and $T^2(f)$ are epi (note $T^2(e)$ is a right inverse for $T^2(f)$) and the fact that θ is a structure map it readily follows that ψ is a structure map. Using 1.1(b) the lemma follows. q.e.d.

Added in proof. The above Lemma (1.6) was independently obtained by Karl Lindblad a student of J.F. Kennison.

§2. Pro-objects and $\mathfrak{M}$-objects

The (dual of the) notion of an $\mathfrak{M}$-object was introduced by Appelgate and Tierney [1] using "atlases". An $\mathfrak{M}$-object was shown to be the same thing as a $\mathbf{T}$-algebra which is a limit of models (by dualizing [1]). A pro-object for $\mathfrak{M}$ was described in [2] pp. 154–166. To within categorical equivalence, the pro-objects can be regarded as the category of all small filtered limits of representable in $(\mathfrak{M}, \mathcal{S})^{\mathrm{op}}$ (that is certain colimits in $(\mathfrak{M}, \mathcal{S})$). We shall regard $\mathfrak{M} \subseteq$ Pro-objects by identifying $M \in \mathfrak{M}$ with the corresponding representable $\mathfrak{M}(M, \ \ ) \in (\mathfrak{M}, \mathcal{S})^{\mathrm{op}}$. For emphasis limits in the category of pro-objects shall be denoted by "pro-lim". If $\{M_i\}$ is filtered then each map pro-lim $\{M_i\} \to M$ must factor through a projection as pro-lim $\{M_i\} \to M_i \to M$. Moreover suppose that the diagram $\{M_i\}$ arises explicitly as $\Delta : D \to M$ (where $M_i = \Delta(i)$). Suppose that f : pro-lim $\Delta \to M$ has two factorizations $f = f_1 p_i$ and $f = f_2 p_j$ through projections p_i and p_j. Then there exists $k \in D$ and $d_1 : k \to i$, $d_1 : k \to j$ in D such that $f_1(d_1) = f_2(d_2)$. (Conversely this condition implies $f_1 p_i = f_2 p_j$.) From this it is possible to obtain a purely formal definition of the pro-objects as formed filtered limits (see [2]).

We wish to know when the category of $\mathfrak{M}$-objects is equivalent to the category of pro-objects for $\mathfrak{M}$. In this section we show that this question is closely related to the question of when Q preserves limits of models. In the next section we give some conditions which imply that Q preserves limits of models.

We proceed to discuss pro-objects and to introduce some definitions which are technically useful. Following [1], observe that there is a lifted singular function $\bar{s} : \mathcal{S}^{\mathbf{T}} \to (\mathfrak{M}, \mathcal{S})^{\mathrm{op}}$ given by $\bar{s}(A) = \mathcal{S}^{\mathbf{T}}(A, \ \ )$. This functor generally has a right adjoint $\bar{r}$ and $\bar{r}\bar{s}$ gives rise to the lifted model induced triple $\bar{\mathbf{T}} = (\bar{T}, \bar{\eta}, \bar{\mu})$ on $\mathcal{S}^{\mathbf{T}}$. (Thus $\bar{T}(A) = \lim\ [(A, \bar{I}\ ) \to \mathcal{S}^{\mathbf{T}}]$ etc. $\bar{r}$ exists if this limit exists which is true if $\mathfrak{M}$ is small or if there is a separating triple for $\bar{I}$.) We now rephrase our question as "When is $\bar{s}$ a full embedding of the $\mathfrak{M}$-objects onto the pro-objects or onto some special class of pro-objects?"

In order to get a reasonable answer to this question we would like $\bar{s}$ to at least preserve the models. That is $\bar{s}(M)$ must be canonically equivalent to the representable, $\mathfrak{M}(M, \ -\ )$. This happens precisely when $\bar{I} : \mathfrak{M} \to \mathcal{S}^{\mathbf{T}}$ is full and faithful (which is usually true, for example the existence of a separating triple guarantees it). When $\bar{I}$ is full and faithful we shall identify $\bar{s}(M)$ with $\mathfrak{M}(M, \ -\ )$ for all models M.

2.1. Proposition. *Let $\bar{I}$ be full and faithful. Then the following statements are equivalent:*

(1) *$\bar{s}$ fully and faithfully embeds the category of $\mathfrak{M}$-objects into* Pro-$\mathfrak{M}$ *(the category of pro-objects).*

(2) *Every $\mathfrak{M}$-object N can be represented as N =* $\mathbf{T}$-lim *M_i where $\{M_i\}$ is filtered and $\bar{s}(N)$ =* pro-lim *$\bar{s}(M_i)$ (recall that* $\mathbf{T}$-lim *M_i is the limit of $\{M_i\}$ in $\mathcal{S}^{\mathbf{T}}$ and* pro-lim *$\bar{s}(M_i)$ is the limit in* Pro-$\mathfrak{M}$).

Moreover, either of the above equivalent statements imply that $\bar{\mathbf{T}}$ is idempotent and $\bar{\mathbf{T}}$ reflects $\mathcal{S}^{\mathbf{T}}$ onto the $\mathfrak{M}$-objects.

Proof. (1) ⇒ (2): Assume that s is such an embedding, and let N be an $\mathfrak{M}$-object. Then $s(N) \in \text{Pro-}\mathfrak{M}$ so $s(N) = \text{pro-lim}\, s(M_i)$. Since s is full and faithful it follows that $N = \mathbf{T}\text{-lim}\, M_i$.

(2) ⇒ (1): If X is a $\mathbf{T}$-algebra and $M \in \mathfrak{M}$ then s maps $\mathcal{S}^{\mathbf{T}}(X, M)$ into the set of natural transformations from $\mathfrak{M}(M, \ )$ to $\mathcal{S}^{\mathbf{T}}(X, \ )$. By Yoneda's lemma (in effect) it follows that the action of s on the hom set $\mathcal{S}^{\mathbf{T}}(X, M)$ is one-to-one and onto. Moreover, if N is a limit of models and s preserves this limit then s is still one-to-one and onto on the hom set $\mathcal{S}^{\mathbf{T}}(X, N)$. Thus (2) implies that s is full and faithful and also that s maps each $\mathfrak{M}$-object into a pro-object.
Finally, if N is an $\mathfrak{M}$-object and $N = \mathbf{T}\text{-lim}\, M_i$ with $s(N) = \text{pro-lim}\, s(M_i)$ then every map from N into a model must factor through a projection (as this is true in $\text{Pro-}\mathfrak{M}$). This means that the projections are initial in the comma category (N, I) so that $\bar{T}(N) = \lim (N, I) = \lim M_i = N$. q.e.d.

Remarks. The next proposition will extend 2.1. We must find some technical remarks and definitions concerning different types of limits. Recall that a $\mathbf{T}_0$-algebra M has *d.c.c. for* $\mathbf{T}_0$*-subalgebras* if every strictly descending chain $M_1 \supsetneq M_2 \supsetneq \dots$ of $\mathbf{T}_0$-subalgebras is necessarily finite. This is equivalent to the *minimum condition for* $\mathbf{T}_0$-subalgebras (that every filtered family of $\mathbf{T}_0$-subalgebras has a smallest member).

Remark (1). If $\mathbf{T}_0$ is a separating triple for $\mathfrak{M}$ and if the $\mathfrak{M}$-objects = pro-objects (via s) then each $M \in \mathfrak{M}$ has d.c.c. for $\mathbf{T}_0$-subalgebras. For assume $M_1 \supsetneq M_2 \supsetneq \dots$ are $\mathbf{T}_0$-subalgebras of M. Let $M_\infty = \cap M_n$ and let $\langle n \rangle : M_\infty \to M_n$ be the inclusion. We also let $\langle m, n \rangle : M_m \to M_n$ be the inclusion for $m > n$. Then $M_\infty = \lim M_n$ in the category of $\mathfrak{M}$-objects, hence also in the category of pro-objects as s is an equivalence. Therefore the identity map $M_\infty : M_\infty \to M_\infty$ must factor as $r\langle n \rangle$ for some $r : M_n \to M_\infty$ (as this is true in pro-objects.) Clearly $\langle n \rangle r \langle n \rangle = M_n \langle n \rangle$ hence $\langle n \rangle r$ and M_n must be equalized by some map in the diagram, that is $\langle n \rangle r \langle m, n \rangle = M_n \langle m, n \rangle$ for some $m > n$ (as this happens in pro-objects). This implies $M_m = M_\infty$.

We can avoid the d.c.c. for subalgebras if we wish to establish conditions for the weaker statement that $\mathfrak{M}$-objects = Regular pro-objects (via s). Regular pro-objects are defined below and from the largest possible class of pro-objects in which $\mathfrak{M}$-objects can be embedded. As we shall show, given d.c.c. for subalgebras, all pro-objects are regular.

Definition. Let $I : \mathfrak{M} \to \mathcal{S}$ be given. Let D be a filtered category. A diagram $\Delta : D \to \mathfrak{M}$ is *regular* if (in addition to being filtered) $\Delta(d)$ is onto for all morphisms d of D (that is $I\Delta(d)$ is onto for all such d). A *regular limit* is the limit of a regular diagram. We are interested in regular limits in $\mathcal{S}^{\mathbf{T}}$ and also in pro-objects. A *regular pro-object* is a regular limit of representables (i.e. we regard $\mathfrak{M} \subset (\mathfrak{M}, \mathcal{S})^{\text{op}}$).

Remark (2). If $\Delta : D \to \mathfrak{M}$ is regular then we may as well assume that D is partially ordered with $i \leq j$ iff there exists $d : i \to j$. Thus if we were to take the dual of D and

allow Δ to become contravariant, then D^{op} is directed and Δ would be a classical inverse limit system all of whose maps are onto. But we do *not* do this. All functors and diagrams are *covariant* here. To prove the assertion note that if $d, d' : i \to j$ are morphisms of D then there exists e with $de = d'e$ hence $\Delta(d) = \Delta(d')$ as $\Delta(e)$ is onto. We can thus identify d and d' without affecting the value of $\lim \Delta$ in $\mathcal{S}^{\mathbf{T}}$ or in pro-objects.

Remark (3). Let $\mathbf{T}_0$ be a separating triple and let every $M \in \mathcal{M}$ have d.c.c. for $\mathbf{T}_0$-subalgebras. Then every pro-object would be regular. Given a filtered diagram $\Delta : D \to \mathcal{M}$, for each $i \in D$ let $\Delta'(i)$ be the smallest $\mathbf{T}_0$-subalgebra of $\Delta(i)$ which is the image of some model $\Delta(j)$ under some map $\Delta(d)$. Then Δ' is regular and $\lim \Delta = \lim \Delta'$ is obvious for $\mathcal{S}^{\mathbf{T}}$ and not difficult to show for pro-objects.

Remark (4). Let $\mathcal{M}$ have and I preserve finite products and let $\mathbf{T}_0$ be a separating triple. Then every $\mathcal{M}$-object can be represented as a regular limit of models. If $X = \lim \{M_\alpha \mid \alpha \in A\}$, let $M_F = \lim \{M_\alpha \mid \alpha \in F\}$ for each finite subset $F \subseteq A$. Let M'_F be the image of X in M_F. Then $\{M'_F\}$ is a regular diagram in $\mathcal{M}$ with $F_1 < F_2$ iff $F_1 \supseteq F_2$, and the obvious maps. Moreover $X = \lim M'_F$.

We need the following definition for technical reasons

Definition. A regular diagram $\{M_i\}$ of models is $\mathbf{T}$-*regular* if every projection $p_i : X \to M_i$ is onto where $X = \lim M_i$ in $\mathcal{S}^{\mathbf{T}}$. (It obviously suffices to let $X = \lim M_i$ in $\mathcal{S}$ or in $\mathcal{S}^{\mathbf{T}_0}$ if $\mathbf{T}_0$ is a separating triple in order to test for $\mathbf{T}$-regularity.) A $\mathbf{T}$-regular diagram is clearly regular.

Let $I : \mathcal{M} \to \mathcal{S}^{\mathbf{T}}$ be full and faithful. Then a pro-object F is a $\mathbf{T}$-*regular pro-object* if $F = \text{pro-lim}\, s(M_i)$ in $(\mathcal{M}, \mathcal{S})^{\mathrm{op}}$ where $\{M_i\}$ is $\mathbf{T}$-regular.

Remark (5). The reason for the above definition is that, as we show in the proposition below, the $\mathbf{T}$-regular pro-objects are the largest class of pro-objects in which s can embed the $\mathcal{M}$-objects. Thus we have sharpened our comments about regular pro-objects. Also the $\mathbf{T}$-regular pro-objects often have a nice topological representation (see 2.4).

It would be interesting to know for which triples $\mathbf{T}$ does regular imply $\mathbf{T}$-regular. It is definitely not true for all $\mathbf{T}$ as it is not true for the identity triple (i.e. it is not true for sets as shown by Henkin in [10]). It is true in compact spaces and in the linearly compact triples as can be shown by the arguments in Section 3. We do not know if it is true for groups. Collecting our remarks we have

2.2. Proposition. *Assume that $\mathcal{M}$ has and I preserves finite products and that $\mathbf{T}_0$ is a separating triple. Then:*

(a) *If s : $\mathfrak{M}$-objects $\to$ Pro-objects is an embedding of the $\mathfrak{M}$-objects into a full subcategory then the range of s is contained in the* **T**-*regular pro-objects and is precisely the* **T**-*regular pro-objects iff s preserves* **T**-*regular limits.*

(b) *If $\{M_i\}$ is* **T**-*regular and X =* **T**-lim $\{M_i\}$ *then* $s(X)$ = pro-lim $s(M_i)$ *iff every map $X \to M \in \mathfrak{M}$ factors as $X \to M_i \to M$ (through a projection). (If $\{M_i\}$ is filtered this condition is still necessary.)*

(c) *Every $\mathfrak{M}$-object is the limit of a* **T**-*regular diagram.*

(d) *s : $\mathfrak{M}$-objects $\to$ Pro-objects is an equivalence of categories iff the condition in* (b) *holds for all* **T**-*regular diagrams; every regular diagram (of models) is* **T**-*regular and each model has d.c.c. for* $\mathbf{T}_0$-*subalgebras.*

Proof. We prove (b) first. Let $\Delta : D \to \mathfrak{M}$ be **T**-regular and let X = **T**-limΔ. Assume that every map $X \to M \in \mathfrak{M}$ factors as $X \to M_i \to M$ for some $i \in D$, where $M_i = \Delta(i)$ and $X \to M_i$ is the projection. We claim that $s(X)$ takes on the same values (as a functor in $(\mathfrak{M}, \mathcal{S})^{\mathrm{op}}$) as the functor pro-lim $\{s(M_i)\}$. This boils down to showing that if $f_1 p_i = f_2 p_j$ (where $i, j \in D$ and p_i, p_j are projections) then there exists $k \in D$, and map $d_1 : k \to i$, $d_2 : k \to j$ with $f_1 \Delta(d_1) = f_2 \Delta(d_2)$.

But as D is filtered there exist k, d_1, d_2 such that $\Delta(d_1)p_k = p_i$ and $\Delta(d_2)p_k = p_j$. Then $f_1\Delta(d_1)p_k = f_2\Delta(d_1)p_k$ and p_k is epi as Δ is **T**-regular. The necessity in (b) is trivial. (c) is a consequence of the construction in Remark (4). As for (a) let X = **T**-lim M_i and assume $s(X)$ = pro-lim $s(M_i)$ where $\{M_i\}$ is filtered. Let M'_i = image of X under the projection $X \to M_i$. Then $\{M'_i\}$ is **T**-regular and X = **T**-lim $\{M'_i\}$ and s preserves this limit by (b). (The necessary condition for s to preserve $\{M_i\}$ implies the sufficient condition for s to preserve $\{M'_i\}$.) Thus $s(X)$ is a **T**-regular pro-object. By 2.1 the range of s is contained in the **T**-regular pro-objects. If the range of s is the **T**-regular pro-objects and if $\{M_i\}$ is a **T**-regular diagram then $s(\lim M_i)$ is the limit (in the range of s) of $\{s(M_i)\}$ which must be the pro-lim $s(M_i)$ as this pro-limit is in the range of s. The converse follows from (c) and 2.1.

Finally (d) is a consequence of (a), 2.1 and Remarks (1) and (3). q.e.d.

We shall now relate the above notions to the case when Q preserves limits of models. First we need:

2.3. Proposition. *(Consequences of Q preserving limits of models.) Let $\mathfrak{M}$ have and I preserve finite products and let* $\mathbf{T}_0$ *be a separating triple. Assume that Q preserves limits of models. Then:*

(a) **T** *is idempotent and its range, the $\mathfrak{M}$-objects, is thus a reflective subcategory of* $\mathcal{S}^{\mathbf{T}}$.

(b) *An $\mathfrak{M}$-object is precisely a closed* $\mathbf{T}_0$-*subalgebra of a product of models. Each such closed subalgebra is given the relative topology by Q.*

(c) *The restriction $Q|\mathfrak{M}$-objects has φF as left adjoint. This pair of adjoint functions generate* $\widetilde{\mathbf{T}}$ *on* Top. *Thus there is a comparison functor*

$$\tau : \mathfrak{M}\text{-objects} \to \mathrm{Top}^{\widetilde{\mathbf{T}}}$$

Moreover, using φ and τ one can regard:

$$\mathcal{M}\text{-objects} \subset \mathrm{Top}^{\tilde{\mathbf{T}}} \subset \mathcal{S}^{\mathbf{T}}.$$

(d) $Q\varphi = \tilde{U}$.

Proof. (a) Let $X = \lim M_i$ be an $\mathcal{M}$-object and let $\eta: X \to \tilde{T}(X)$ be the unit of $\tilde{\mathbf{T}}$ at X. By construction of $\tilde{T}(X)$ there exists a retraction $\theta: \tilde{T}(X) \to X$. Since η and θ are $\tilde{\mathbf{T}}$-homomorphisms they are Q-continuous hence $\eta(X)$ is closed in the Q-topology on $\tilde{T}(X)$. But $\eta(X)$ is dense in the limit topology (by using the type of argument used in 1.4). By hypothesis, we see $\eta(X) = \tilde{T}(X)$ hence $X = \tilde{T}(X)$ and $\tilde{\mathbf{T}}$ is idempotent. The rest of (a) is obvious.

(b) Let $X \subset \prod\{M_\alpha \mid \alpha \in A\}$ be a closed $\mathbf{T}_0$-subalgebra of a product of models. Let D be the class of all finite subsets $F \subset A$ partially ordered by $F_1 < F_2$ iff $F_1 \supset F_2$. (Then D happens to be filtered.) Given $F \in D$ let M_F be the projection of X on $\prod\{M_\alpha \mid \alpha \in F\}$. Let $L = \mathbf{T}\text{-lim}\,\{M_F\}$ then we can construct L so that $X \subset L \subset \prod M_\alpha$. By the now familiar argument, X is dense in L but X is closed so $X = L$. Thus X is an $\mathcal{M}$-object. (Note that in view of the hypothesis the topologies on L and $\prod M_\alpha$ are unambigious.) The converse is obvious.

(c) Let $X \in \mathrm{Top}$ be given. Then $\tilde{F}(X)$ is a $\tilde{\mathbf{T}}$-limit of models and since $Q\varphi$ preserves limits of models, $Q\varphi\tilde{F}(X) = \tilde{T}(X)$. We claim that $\tilde{\eta}: X \to \tilde{T}(X)$ is a front adjunction. Let $N = \lim M_i$ be an $\mathcal{M}$-object and let $h: X \to Q(N)$ be continuous. By construction of $\tilde{T}(X)$ there exists a unique $\mathbf{T}$-homomorphism $f: \tilde{T}(X) \to N$ such that $f\tilde{\eta} = h$ (by considering projections). It follows from this proof that $\tilde{\mathbf{T}}$ is the triple generated by the adjointness between φF and $Q \mid (\mathcal{M}\text{-objects})$. Hence the comparison function τ exists. The last part of (c) is the statement that φ is a full embedding (which will follow from 1.4(d) and 2.3(d) (proven below)) and that $\varphi\tau$ is the inclusion if $\mathcal{M}$-objects into $\mathcal{S}^{\mathbf{T}}$. But φ and τ both preserve models and both preserve limits (as they are comparison functions). Hence $\varphi\tau$ preserves $\mathcal{M}$-objects.

(d) Let $(X, \theta) \in \mathrm{Top}^{\tilde{\mathbf{T}}}$. Then $\tilde{T}(X)$ has an unambigious topology. $\theta: \tilde{T}(X) \to X$ is a quotient map (onto the $\tilde{U}$ topology) since $\theta\tilde{\eta} = X$. But θ in Q-continuous so every Q-open subset of X is $\tilde{U}$-open. On the other hand $\theta\lambda$ is $\tilde{U}$-continuous, but by definition is a quotient map from $T(X)$ onto $Q(X, \theta\lambda) = Q\varphi(X, \theta)$. Thus a $\tilde{U}$-open subset is Q-open. q.e.d.

Definition. Let $\mathbf{T}_0$ be a finitary triple over $\mathcal{S}$ which admits a group operation. Then the notion of a complete topological $\mathbf{T}_0$*-algebra* is clear (using the usual notion of a complete topological group).

If $\mathcal{M}$ is any collection of (discrete) $\mathbf{T}_0$-algebras we say that the topological $\mathbf{T}_0$-algebra, X, is *$\mathcal{M}$-generated* if X has a basic system of neighbourhoods $\{U_\alpha\}$ at 1 such that each U_α is an open $\mathbf{T}_0$-kernel and X/U_α is algebraically equivalent to a member of $\mathcal{M}$.

2.4. Lemma. *Let $\mathfrak{M}$ and $\mathbf{T}_0$ be as in the above definition and assume that $\mathfrak{M}$ has and I preserves finite products. Then the category of $\mathbf{T}$-regular pro-objects is equivalent to the category of $\mathbf{T}$-regular limits (in the category of topological $\mathbf{T}_0$-algebras) of discrete models. This is precisely the category of complete, Hausdorff, $\mathfrak{M}$-generated topological $\mathbf{T}_0$-algebras.*

Proof. Let $\{M_i\}$ be a $\mathbf{T}$-regular diagram of models. Let $\lim M_i$ be the limit in topological $\mathbf{T}_0$-algebras (so that it has the limit topology). Let $f: \lim M_i \to M$ be a continuous $\mathbf{T}_0$-homomorphism. Then $\operatorname{Ker} f$ is open so there exists a projection p_i with $\operatorname{Ker} p_i \subseteq \operatorname{Ker} f$ (as $\{\operatorname{Ker} p_i\}$ is a base at 1). Thus f factors through p_i as p_i is onto. The proof of 2.2(b) now applies.

If X is complete and Hausdorff and has a base $\{U_\alpha\}$ of open $\mathbf{T}_0$-kernels with $X/U_\alpha \in \mathfrak{M}$ then $\{X/U_\alpha\}$ is a $\mathbf{T}$-regular diagram whose limit is X. q.e.d.

2.5. Proposition. *Assume that $\mathfrak{M}$ has and I preserves finite products. Let $\mathbf{T}_0$ be a finitary separating triple which admits a group operation. Further, for convenience (see 1.5) assume that the models are closed under the formation of $\mathbf{T}_0$-quotients. Then:*

(a) *The $\mathbf{T}$-regular pro-objects is equivalent to the category of complete $\tilde{\mathbf{T}}$-algebras.*

(b) *If Q preserves limits of models, then the following five categories are canonically equivalent (see 2.2, 2.3(c) and (a) above):*

$\mathfrak{M}$-objects = complete $\tilde{\mathbf{T}}$-algebras
= complete, Hausdorff $\mathbf{T}$-algebras
= $\mathbf{T}$-regular pro-objects
= complete, Hausdorff, $\mathfrak{M}$-generated $\mathbf{T}_0$-algebras.

(c) *Conversely, if $\mathfrak{M}$-objects = $\mathbf{T}$-regular pro-objects (via s) then Q preserves limits of models.*

Proof. (a) The proof of 1.5(b) can be applied to show that all $\tilde{\mathbf{T}}$-algebras are $\mathfrak{M}$-generated (the additional use of 1.3(e) makes this easier to see). By 1.3(e) and 2.4 and the observation that all models are $\tilde{\mathbf{T}}$-algebras and that $\operatorname{Top}^{\tilde{\mathbf{T}}}$ has all limits, which are preserved by $\tilde{U}$, (a) follows.

(b) It easily follows from 2.4, 1.3(b) and (a) above, that all five of these categories are equivalent to the category of complete, Hausdorff, $\mathfrak{M}$-generated topological $\mathbf{T}_0$-algebras. That these equivalence are "canonical" follows from noting that all of the equivalence preserve the models and limits and the topology (by 2.3(d) and the fact that Q and $\tilde{U}$ preserve topological limits of models). That the $\mathfrak{M}$-objects are equivalent to the $\mathbf{T}$-regular pro-object *via* s follows from 2.3(a), 2.2 and the proof of 2.4.

(c) Conversely, assume that the $\mathfrak{M}$-objects are equivalent to the $\mathbf{T}$-regular pro-objects via s. Then the condition in 2.2(b) holds for every $\mathbf{T}$-regular diagram.

Now let $X = \mathbf{T}\text{-}\lim \{M_i\}$. Claim that $Q(X) = \operatorname{Top}\text{-}\lim \{M_i\}$. We may as well assume that $\{M_i\}$ is $\mathbf{T}$-regular (as we can replace $\{M_i\}$ by the diagram $\{M'_F\}$ used in Remark (4) without changing X or the limit topology. By construction $\{M'_F\}$ is $\mathbf{T}$-regular).

By 1.5(b), $Q(X)$ is the smallest topology rendering each $\mathbf{T}$-homomorphism $X \to M$ continuous. By assumption each such $X \to M$ factors as $X \to M_i \to M$ where $X \to M_i$ is a projection. Hence Top-lin$\{M_i\}$renders each such $X \to M$ continuous and the claim follows (e.g. by 1.1(b)). q.e.d

§3. The compact and linearly compact cases

Definition. Following Manes [16, p. 104] we say that a full subcategory $B \subseteq \mathcal{S}^{\mathbf{T}}$ is a *Birkhoff subcategory of* $\mathcal{S}^{\mathbf{T}}$ if B is closed under the formation of products, $\mathbf{T}$-subalgebras and $\mathbf{T}$-quotients. Manes proves that Birkhoff subcategories are tripleable over $\mathcal{S}$. Conversely, if $\mathcal{S}^{\mathbf{T}'}$ is a full tripleable subcategory of $\mathcal{S}^{\mathbf{T}}$ and if the induced map $T(X) \to T'(X)$ is onto for all X then $\mathcal{S}^{\mathbf{T}'}$ is a Birkhoff subcategory. From an equational point of view, a Birkhoff subcategory is determined by equational identities (so that the Birkhoff subcategory is the class of all algebras satisfying the identities). In the category of rings, for example, the subcategory of all commutative rings is the Birkhoff subcategory determined by the identity $xy = yx$.

3.1. Theorem *(the compact case). Let $\mathcal{M}$ have and I preserve finite products. Let $\mathbf{T}_0$ be a finitary separating triple and assume that every model is finite. Then $\mathcal{S}^{\mathbf{T}}$ is the smallest Birkhoff subcategory (containing the models) of the compact, Hausdorff $\mathbf{T}_0$-algebras. Moreover Q preserves limits and $Q\varphi = U$ hence φ is a full embedding. Finally:*

$$\mathcal{M}\text{-objects} = \text{Pro-objects} \subseteq \mathrm{Top}^{\widetilde{\mathbf{T}}} \subseteq \mathcal{S}^{\mathbf{T}}.$$

If $\mathbf{T}_0$ admits a group operation these categories coincide.

Proof. Since the category C of compact Hausdorff $\mathbf{T}_0$-algebras is equational (see [16]) and contains the models, there exists a limit preserving forgetful functor $V: \mathcal{S}^{\mathbf{T}} \to C$ (by definition of the equational completion or since every operation of C is an operation of $\mathcal{S}^{\mathbf{T}}$). If $(X, \theta) \in \mathcal{S}^{\mathbf{T}}$ then $V(\theta): VT(X) \to VX$ is a closed mapping hence a quotient map so the C-topology on VX is precisely $Q(X, \theta)$. Thus, by 1.3(b), V is a full embedding. Moreover if $F_C: \mathcal{S} \to C$ is the free functor then by 1.4(c) the canonical map $F_C(X) \to T(X)$ is dense hence onto. Thus $\mathcal{S}^{\mathbf{T}}$ is a Birkhoff subcategory and clearly the smallest one containing the models by the fact that $\mathcal{S}^{\mathbf{T}}$ is an equational completion. By 1.1(b) the Q-topology on a $\mathbf{T}$-limit is at least as big as the topological limit but both are compact, Hausdorff hence they coincide. Thus Q preserves limit so $Q\varphi = U$ by 2.3(d) and φ is a full embedding by 1.3(d).

It remains to show that $\mathcal{M}$-objects = Pro-objects in view of 2.3(c). (The statement concerning the case when $\mathbf{T}_0$ has a group operation is a special case of Theorem 3.3 proved below.) Using 2.2(b) let M_i be $\mathbf{T}$-regular and let $X = \mathbf{T}\text{-lim}\, M_i$. Let $f: X \to M$ be given. Each $x \in X$ has a neighbourhood of the form $p_i^{-1}(m_i)$ on which f is constant as M is discrete. Using compactness and filteredness there exists a pro-

jection p_i such that f is constant on all sets of the form $p_i^{-1}(m)$. Since p_i is onto, f factors through p_i by considering congruence relations.

We shall pursue 2.2(d). Note that $\bar{\mathbf{T}}$ is idempotent by 2.3(a) and d.c.c. for $\mathbf{T}_0$-subalgebras of models is trivial as all models are finite. Thus it suffices to show that if M_i is regular then it is $\mathbf{T}$-regular. By a previous remark we may as well assume that i varies in a filtered, partially ordered set. Let $X = \mathbf{T}\text{-lim}\, M_i$ and let $p_i : X \to M_i$ be the projection. Regard $X \subseteq \prod M_i$ and let $p_i : \prod M_i \to M_i$ be the projection for the product. For each $d : M_i \to M_k$ in the diagram, let:

$$A_d = \{y \in \textstyle\prod M_i ; d\, p_i(y) = p_k(y)\}.$$

Then A_d is closed and we can regard $X = \cap A_d$. Let i_0 and $m \in M_{i_0}$ be chosen. Then $(p_{i_0})^{-1}(m) \cap A_d$ has f.i.p. (as can readily be shown by using the fact that the diagram is filtered). By compactness of $\prod M_i$ there is an x in the intersection so $x \in \cap A_d = X$ and $p_{i_0}(x) = m$. Hence the projections are onto. q.e.d.

Definition. Let $\mathbf{T}_0$ be a finitary triple over sets which admits a group operation. Then a $\mathbf{T}_0$-*kernel coset* is a set of the form Ux where U is a $\mathbf{T}_0$-kernel. The notion of a *topological* $\mathbf{T}_0$-*algebra* is well known. A topological $\mathbf{T}_0$-algebra is *linearly compact* if every family of closed $\mathbf{T}_0$-kernel cosets with f.i.p. (finite intersection property) has non-void intersection.

Notice that a discrete $\mathbf{T}_0$-algebra is linearly compact if it has d.c.c. for $\mathbf{T}_0$-kernels. The converse is not true (see [18]). The term "pseudocompact" is used for a special case of linear compactness in [5].

Definition. $\mathbf{T}_0$ is a *normal separating triple* for $I : \mathcal{M} \to \mathcal{S}$ if $\mathbf{T}_0$ is a finitary triple with a group operation; the models are closed under $\mathbf{T}_0$-quotients (see 1.5); and $\mathcal{M}$ has and I preserves finite products.

$I : \mathcal{M} \to \mathcal{S}$ satisfies the *linearly compact conditions* (abreviated LCC) with respect to $\mathbf{T}_0$ if $\mathbf{T}_0$ is a normal separating triple and each $M \in \mathcal{M}$ is linearly compact in the discrete topology.

Definition. Let $\mathbf{T}_0$ admit a group operation. Let $\mathcal{M}$ be a given class of (discrete) $\mathbf{T}_0$-algebras. Let X be a topological $\mathbf{T}_0$-algebra. Then X is *strongly* $\mathcal{M}$-*generated* if it is $\mathcal{M}$-generated and if $K \subseteq X$ is open whenever K is a closed $\mathbf{T}_0$-kernel with X/K algebraically equivalent to some $M \in \mathcal{M}$.

Remarks. Linearly compact modules have often been examined in the literature (e.g. see [5, 12, 13, 18], however their tripleability properties over $\mathcal{S}$ and sometimes Top seem not to have been established before). It is convenient here to list some of the elementary results that extend to the general case of $\mathbf{T}_0$-algebras and which are useful below or in considering examples.

(1) Discrete linearly compact $\mathbf{T}_0$-algebras are closed under the formation of finite products and $\mathbf{T}_0$-quotients, but not necessarily under the formation of $\mathbf{T}_0$-subalgebras. (In fact, infinite nested intersections of linearly compact $\mathbf{T}_0$-subalgebras need not be linearly compact.)

(2) Hausdorff linearly compact $\mathbf{T}_0$-algebras having a base of open $\mathbf{T}_0$-kernels at 1 are closed under the formation of regular limits and arbitrary products and $\mathbf{T}_0$-quotients (by closed $\mathbf{T}_0$-kernels).

(3) If $\mathcal{M}$ is any full subcategory of discrete linearly compact $\mathbf{T}_0$-algebras then, for $\mathcal{M}$-generated, Hausdorff topological $\mathbf{T}_0$-algebras, completeness is equivalent to linear compactness. Also the complete (or linearly compact) Hausdorff $\mathcal{M}$-generated topological $\mathbf{T}_0$-algebras are category equivalent to the regular pro-objects for $\mathcal{M}$.

These remarks can easily be proved either by generalizing the proofs in [12] and [18] or by using the methods in this section. We shall however sketch the proofs. If A and B are discrete, linearly compact $\mathbf{T}_0$-algebras, let $\mathcal{F}$ be a maximal filter of $\mathbf{T}_0$-kernels cosets of $A \times B$. Let $\mathcal{F}_A = \{p_A(F) \mid F \in \mathcal{F}\}$. There exists $a \in \cap\, \mathcal{F}_A$ hence $p_A^{-1}(a) \in \mathcal{F}$ as it meets all members of $\mathcal{F}$ and $\mathcal{F}$ is maximal. Similarly there exists $b \in B$ with $p_B^{-1}(b) \in \mathcal{F}$ so $(a, b) \in \cap\, \mathcal{F}$. Thus $A \times B$ is linearly compact. A similar argument (replacing $p_A(F)$ by its closure) can be used to prove (2). Alternatively one can generalize the proof of 3.13. Most of (3) follows from 2.4 (that regular limits are $\mathbf{T}$-regular is essentially known and also proved as 3.11 below). That linear compactness implies completeness follows since maximal filters containing small sets have subfilters of closed $\mathbf{T}_0$-kernel cosets.

Remark. Several of the other interesting properties for linear compactness (cf. [18]) can be generalized to $\mathbf{T}_0$-algebras. Also some of the lemmas below such as 3.9, 3.10, 3.12 and 3.14 can be generalized to Hausdorff linearly compact $\mathbf{T}_0$-algebras having a base of open $\mathbf{T}_0$-kernels.

3.2. Theorem *(the linearly compact case). Assume that $I : \mathcal{M} \to \mathcal{S}$ satisfies LCC with respect to $\mathbf{T}_0$. Then:*

(a) $\mathcal{S}^{\mathbf{T}}$ *is equivalent (using the Q-topology) to the category of strongly $\mathcal{M}$-generated complete, Hausdorff $\mathbf{T}_0$-algebras. Every member of $\mathcal{S}^{\mathbf{T}}$ is linearly compact. Moreover $\mathbf{T}$ is the identity triple on $\mathcal{S}^{\mathbf{T}}$ so that $\mathcal{S}^{\mathbf{T}} = \tilde{\mathcal{M}}$-objects.*

(b) $\mathrm{Top}^{\tilde{\mathbf{T}}}$ = *Regular Pro-objects. Alternatively, $\mathrm{Top}^{\tilde{\mathbf{T}}}$ is equivalent (using the $\tilde{U}$-topology) to the category of $\mathcal{M}$-generated complete Hausdorff $\mathbf{T}_0$-algebras. Every member of $\mathrm{Top}^{\tilde{\mathbf{T}}}$ is linearly compact. Note that we can regard $\mathcal{S}^{\mathbf{T}} \subset \mathrm{Top}^{\tilde{\mathbf{T}}} \subset$ Top $\mathbf{T}_0$-algebras. In this case $\varphi : \mathrm{Top}^{\tilde{\mathbf{T}}} \to \mathcal{S}^{\mathbf{T}}$ is a coreflective functor (that is the right adjoint of the inclusion $\mathcal{S}^{\mathbf{T}} \to \mathrm{Top}^{\tilde{\mathbf{T}}}$). Explicitly φ preserves the underlying $\mathbf{T}_0$-algebra structure and simply increases the number of open sets. If $X \in \mathrm{Top}^{\tilde{\mathbf{T}}}$ then*

the set of all closed $\mathbf{T}_0$*-kernels* K *such that* X/K *is algebraically a model becomes a base at* 1 *for* $\varphi(X)$. *The case when* φ *is a category equivalence is discussed in* 3.3.

(c) A *is a* $\mathbf{T}$*-subalgebra of* $X \in \mathcal{S}^{\mathbf{T}}$ *iff* A *is a closed* $\mathbf{T}_0$*-subalgebra (in the* Q*-topology).* ($Q(A)$ *may presumably fail to be the relative topology, but see* 3.3.)

3.3. Theorem *(the d.c.c. case). Let* $\mathbf{T}_0$ *be a normal separating triple for* $I : \mathcal{M} \to \mathcal{S}$. *Assume that each model has d.c.c. with respect to* $\mathbf{T}_0$*-kernels (thus LCC is satisfied). Then (in addition to the results of* 3.2):

(a) φ *is a category equivalence.*

(b) Q *preserves all limits.*

(c) *If* $A \subseteq X$ *is a* $\mathbf{T}$*-subalgebra then* $Q(A)$ *has the relative topology.*

(d) $Q\varphi = \tilde{U}$. *(Thus* $\mathcal{S}^{\mathbf{T}}$ *is tripleable over* Top *via* Q.)

(e) *If* $f : A \to B$ *is an onto* $\mathbf{T}$*-homomorphism then* f *admits a continuous section (that is* $Q(f)$ *is a split epi).* This was proved by Serre [17] for pro-finite groups and by Brumer [5], for pseudocompact modules (see Example 4.3).

Moreover if $\mathbf{T}_0$ *is a normal separating triple, then* (b) *and* (c) *together imply the d.c.c. assumption on* $\mathbf{T}_0$*-kernels. This d.c.c. assumption is, in the presence of LCC equivalent to* (a), (b) *and* (d), *individually.*

The proofs of 3.2 and 3.3 shall be postponed until we establish some lemmas concerning certain regular limits of models. For technical reasons we set up the following notation and definitions.

Notation. In what follows D shall always denote a small, filtered, partially ordered category and $\Delta : D \to \mathcal{M}$ shall be regular. If $g \in D$ then we generally let $M_g = \Delta(g)$ and if $g < h$ in D then the induced map $M_g \to M_h$ shall be denoted as $\delta(g, h)$ or just δ if there is no danger of confusion.

The canonical projection $\lim\Delta \to M_g$ shall be denoted by $\langle g \rangle$. This extends the notation for the important case where $D = (X, I)_0$ and Δ is chosen so that $\lim\Delta = T(X)$ etc.

Definition. Let $\Delta : D \to \mathcal{M}$ be as above. Given $g, h \in D$ and $a \in M_h$ we choose $k < g, h$ and let $\delta_1 = \delta(k, h)$ and $\delta_2 = \delta(k, g)$. We define $\tau_{g,h}(a) = \delta_2(\delta_1^{-1}(a))$. Then $\tau_{g,h}(a) \subseteq M_g$ is independent of the choice of k. (For $b \in M_g$ is in $\tau_{g,h}(a)$ iff there exists $x \in M_k$ with $\delta_1(x) = a$ and $\delta_2(x) = b$. If $k' < k$ then there exists such an $x \in M_k$ iff there exists a suitable $x' \in M_{k'}$ as $\delta : M_{k'} \to M_k$ is onto. Since D is filtered, given k_1 and k_2 there exists $k' < k_1, k_2$ and the independence of the choice of k is obvious.)

We note that if $\zeta \in \lim\Delta$ and $\langle h \rangle(\zeta) = a$ and $\langle g \rangle(\zeta) = b$ then $b \in \tau_{g,h}(a)$. We shall establish a general result which proves the converse.

Definition. Let $\Delta : D \to \mathcal{M}$ be as above. Then (H, γ) is a Δ*-prescription* (or simply a

prescription) if $H \subseteq D$ and if γ is a function such that $\gamma(h) \in M_h$ for all $h \in H$. We call (H_0, γ_0) a *subprescription* of (H, γ) [and (H, γ) an ***extension*** of (H_0, γ_0)] if $H_0 \subseteq H$ and $\gamma | H_0 = \gamma_0$. A *solution* of (H, γ) is a point $\zeta \in \lim \Delta$ such that $\langle h \rangle(\zeta) = \gamma(h)$ for all $h \in H$.

(H, γ) is *consistent* if it has a solution and is *finitely consistent* if every finite subprescription is consistent.

Definition. Let (H, γ) be a Δ-prescription and let $g \in D$ be given. Then we define

$$\hat{\gamma}(g) = \cap \{\tau_{g,h}(a) \mid h \in H, a = \gamma(h)\}.$$

We say that (H, γ) is *formally consistent* if $\hat{\gamma}(g) \neq \emptyset$ for all $g \in D$. We call (H, γ) *finitely formally consistent* if every finite subprescription is formally consistent. Observe that a (finitely) consistent prescription in (finitely) formally consistent.

Definition. $\Delta : D \to \mathfrak{M}$ is *very regular* if (Δ is regular and) D has infs (denoted $g \wedge h$) and the induced map $\Delta(g \wedge h) \to \Delta(g) \times \Delta(h)$ is one-one. (For example if $D = (X, I)_0$ and $T(X) = \lim \Delta$ etc. then Δ is very regular, see discussion preceeding 1.3 assuming there exists a well-behaved $\mathbf{T}_0$.)

Definition. Let $\Delta : D \to \mathfrak{M}$ be very regular. Then the Δ-prescription (H, γ) is $\wedge$-*closed* if $h_1 \wedge h_2 \in H$ whenever $h_1, h_2 \in H$ and $\gamma(h_1 \wedge h_2)$ is the unique element of $\Delta(h_1 \wedge h_2)$ which maps into $(\gamma(h_1), \gamma(h_2)) \in \Delta(h_1) \times \Delta(h_2)$.

Most of the following lemmas have straightforward proofs which are omitted or sketched.

3.4. Lemma. *Let $\Delta : D \to \mathfrak{M}$ be very regular. If (H, γ) is $\wedge$-closed then (H, γ) is finitely formally consistent.*

3.5. Lemma. *Given LCC, a finitely formally consistent diagram is formally consistent.* (Note $\tau_{g,h}(a)$ is a $\mathbf{T}_0$-kernel coset by applying 1.4.)

3.6. Lemma. *Let $f, g, h \in D$ with $f \leq g$. Let $a \in M_h$. Then $\delta : M_f \to M_g$ maps $\tau_{f,h}(a)$ onto $\tau_{g,h}(a)$. Therefore if (H, γ) is a prescription and $f \leq g$ then δ map $\hat{\gamma}(f)$ into $\hat{\gamma}(g)$.*

3.7. Lemma. *Assume LCC and that $\Delta : D \to \mathfrak{M}$ is very regular. Let (H, γ) be $\wedge$-closed and let $f \leq g$. Then δ map $\hat{\gamma}(f)$ onto $\hat{\gamma}(g)$.*

Proof. Given $s \in \hat{\gamma}(g)$ then $\delta^{-1}(s) \cap \tau_{f,h}(a)$ (where $h \in H$ and $a = \gamma(h)$) has f.i.p. (by 3.6 and as H has finite infs). Apply LCC. q.e.d.

3.8. Lemma. *Assume LCC and that $\Delta : D \to \mathfrak{M}$ is very regular. Let (H, γ) be $\wedge$-closed. Let $g \in D$ and $b \in \hat{\gamma}(g)$ be given. Let $G = \{g \wedge h \mid h \in H\} \cup \{g\} \cup H$. Then there exists a unique β such that (G, β) is a $\wedge$-closed extension of (H, γ) and $\beta(g) = b$.*

Proof. Given $h \in H$ let $\delta_1 : \Delta(g \wedge h) \to \Delta(g)$ and $\delta_2 : \Delta(g \wedge h) \to \Delta(h)$. By 3.7 there exists $x \in \hat{\gamma}(g \wedge h)$ such that $\delta_1(x) = b$. Then $\delta_2(x)$ must be $\gamma(h)$ (by 3.6) and x is uniquely determined as Δ is very regular. Define $\beta(g \wedge h) = x$. Since x is determined by $\delta_1(x) = b$ we see that this definition is unambigious in case $g \wedge h = g \wedge (h')$. By the same argument if we further define $\beta(g) = b$ and $\beta(h) = \gamma(h)$ for $h \in H$ we see that β is still well-defined and (G, β) is $\wedge$-closed.

3.9. Lemma. *Assume LCC and that $\Delta : D \to \mathfrak{M}$ is very regular. Then every $\wedge$-closed prescription is consistent.*

Proof. Let (H, γ) be $\wedge$-closed. By Zorn's Lemma there exists a maximal $\wedge$-closed extension (H^*, γ^*). By 3.8 we seen that $H^* = D$ hence γ^* defines a member of $\lim \Delta$ which is a solution of (H, γ). q.e.d.

3.10. Lemma. *Assume LCC (and let $\Delta : D \to \mathfrak{M}$ be regular). Then every finitely consistent Δ-prescription is consistent.*

Proof. Given $h_1, \ldots, h_n \in D$ choose $f \leq h_i$ for $i = 1, \ldots n$ and let $\Delta(f) \to \Delta(h_1) \times \ldots \times \Delta(h_n)$ be the obvious map. Then the image of $\Delta(f)$ is independent of the choice of f (for the same reason that $\tau_{g,h}(a)$ is independent of the choice of k).

We now extend D to D' by adjoining formal finite infs to D and extend Δ to Δ' so that $\Delta'(h_1 \wedge \ldots \wedge h_n)$ is the above defined image of $\Delta(f)$. Then D is initial in D' so $\lim \Delta = \lim \Delta'$ and $\Delta' : D' \to \mathfrak{M}$ is very regular by construction.

If (H, γ) is a finitely consistent Δ-prescription then (H, γ) can readily be extended to a $\wedge$-closed Δ'-prescription which is consistent by 3.9. Thus (H, γ) is consistent.

q.e.d.

3.11. Corollary. *Assume LCC. Then every regular diagram is* **T**-*regular.*

Proof. Using Remark (2), preceeding 2.2, it suffices to show that the regular diagram $\Delta : D \to \mathfrak{M}$ is **T**-regular where D is partially ordered. Let $h \in D$ and $a \in M_h$ be given. Let $H = \{h\}$ and $a = \gamma(h)$. Then (H, γ) is a $\wedge$-closed prescription for the extended diagram $\Delta' : D' \to \mathfrak{M}$ constructed above. By 3.9 there exists $\zeta \in \lim \Delta'$ such that $\langle h \rangle(\zeta) = a$ so Δ is **T**-regular. q.e.d.

Definition. Given LCC then the *$T(X)$-diagram* (for $X \in \mathfrak{J}$) shall refer to the canonical functor $(X, I)_0 \to \mathfrak{M}$. This diagram is very regular (see the discussion proceding 1.3). A prescription for this diagram shall be called a *$T(X)$-prescription*.

3.12. Proposition. *Assume LCC. Let $(X, \theta) \in \mathcal{S}^{\mathbf{T}}$ and let H be a set of $\mathbf{T}$-homomorphisms from X into models. For each $h: X \to M$ in H let $\gamma(h) \in M$. If (H, γ) is finitely consistent (i.e. given $H_0 \subseteq H$, a finite subset, there exists $x \in X$ with $h(x) = \gamma(h)$ for all $h \in H_0$) then (H, γ) is consistent (i.e. there exists $x \in X$ with $h(x) = \gamma(h)$ for all $h \in H$).*

Proof. We may as well assume that each $h \in H$ is onto. Then $H \subseteq (X, I)_0$ and (H, γ) may be regarded as a $T(X)$-prescription. If $x \in X$ is such that $h(x) = \gamma(h)$ for all $h \in H_0$ then $\langle h\rangle(\eta(x)) = \gamma(h)$ for all $h \in H_0$ so (H, γ) is obviously finitely consistent. By 3.10 there exists $\zeta \in T(X)$ with $\langle h\rangle(\zeta) = \gamma(h)$ for all $h \in H$. Then $x = \theta(\zeta)$ has the desired property. (Note $\langle h\rangle = h\theta$ since h is a $\mathbf{T}$-homomorphism for all $h \in H$.) q.e.d.

3.13. Proposition. *Assume LCC. Let $(X, \theta) \in \mathcal{S}^{\mathbf{T}}$. Then (using the Q-topology) (X, θ) is a linearly compact $\mathbf{T}_0$-algebra. It follows that (X, θ) is complete from the remark preceeding Theorem 3.2.*

Proof. By 1.2, (X, θ) with the Q-topology can be regarded as a topological $\mathbf{T}_0$-algebra. Let $\{U_i\}$ be a collection of closed $\mathbf{T}_0$-kernels of X and assume that the cosets $\{U_i x_i\}$ have f.i.p. We must show $\cap\{U_i x_i\} \neq \emptyset$. Since each closed $\mathbf{T}_0$-kernel is an intersection of open $\mathbf{T}_0$-kernels (see 1.4 which applies in view of 1.5(b)) it clearly suffices to assume that each U_i is an open $\mathbf{T}_0$-kernel. By 1.5(b) it follows that $X/U_i = M_i$ is a model. Let $h_i: X \to M_i$ be the corresponding projection. Let $H = \{h_i\}$ and $\gamma(h_i) = h_i(x_i)$. Then (H, γ) satisfies the hypotheses of 3.12 as finite consistency is here equivalent to the f.i.p. condition. The conclusion of 3.12 implies $\cap\{U_i x_i\} \neq \emptyset$. q.e.d.

3.14. Lemma. *Assume LCC. Let (X, θ) and (Y, ψ) be $\mathbf{T}$-algebras with $(X, \theta) = \mathbf{T}\text{-}\lim M_i$ (a limit of models). Let $g: (X, \theta) \to (Y, \psi)$ be a $\mathbf{T}$-homomorphism. Then $\operatorname{Ker} g$ is closed in the limit topology on X.*

Proof. We first claim that $\operatorname{Ker} g = \theta(\operatorname{Ker} Tg)$. For if $\zeta \in \operatorname{Ker} Tg$ then $g(\theta\zeta) = \psi(Tg)(\zeta) = 1$ hence $\theta(\zeta) \in \operatorname{Ker} g$. Conversely, let $x \in \operatorname{Ker} g$. Then $x = \theta(\eta x(\eta 1)^{-1})$ and $\eta x(\eta 1)^{-1} \in \operatorname{Ker} Tg$ as $(Tg)\eta x = \eta_Y(gx)$ and $(Tg)(\eta 1)^{-1} = [\eta_Y(g1)]^{-1}$ and $gx = g1 = 1$.

We next claim that $\theta(\operatorname{Ker} Tg)$ is closed in the limit topology. Assume that t is in the closure of $\theta(\operatorname{Ker} Tg)$ in the limit topology. Then for any finite set of indices say $\{1, 2, \ldots, n\}$ (so labelled for convenience), there exists $\zeta \in \operatorname{Ker} Tg$ with $p_i\theta(\zeta) = p_i(t)$ for $i = 1, \ldots, n$ (where $p_i: X \to M_i$ is the projection). But $p_i\theta(\zeta) = \langle p_i\rangle(\zeta)$ as p_i is admissible and $\zeta \in \operatorname{Ker} Tg$ iff $\langle f\rangle(Tg)(\zeta) = \langle gh\rangle(\zeta) = 1$ for all $f: Y \to M$. By applying 3.12 to $T(X)$ we can find $\zeta \in TX$ such that $\langle p_i\rangle(\zeta) = p_i(t)$ for all i and $\langle gf\rangle(\zeta) = 1$ for all $f: Y \to M$. Then $\theta(\zeta) = t$ and $\zeta \in \operatorname{Ker} Tg$ so $t \in \theta(\operatorname{Ker} Tg)$. q.e.d.

3.15. Corollary. *Assume LCC. Let $(X, \theta) \in \mathcal{S}^{\mathbf{T}}$. Then $Q(X, \theta)$ is Hausdorff.*

Proof. $\theta : TX \to X$ is a **T**-homomorphism and TX is a limit of models hence Ker θ is closed in TX. Thus $1 \in X$ is closed in the quotient topology, $Q(X, \theta)$, which implies $Q(X, \theta)$ is Hausdorff as X is a topological group. q.e.d.

3.16. Corollary. *Assume LCC. Then* $\mathbf{T}$ *is idempotent and trivial (that is T is the identity functor). Thus* $\mathcal{S}^{\mathbf{T}} = \mathfrak{M}$*-objects and every* $X \in \mathcal{S}^{\mathbf{T}}$ *is a canonical limit of models.*

Proof. Let $X \in \mathcal{S}^{\mathbf{T}}$. Then $\eta : X \to T(X)$ is one-one in view of 1.5(b) since $Q(X)$ is Hausdorff. But by applying 3.11 and by representing $T(X)$ as a regular filtered limit of models we see that $\eta(X)$ is dense in $T(X)$ (when $T(X)$ is given the limit topology which need not be $Q\overline{T}(X)$). By 3.13, X is complete hence is closed in the limit topology on $T(X)$. Thus η is onto, hence is an equivalence. q.e.d.

Proof of Theorem 3.2. (a) Let $U_0 : \mathcal{S}^{\mathbf{T}} \to \mathcal{S}^{\mathbf{T}_0}$ be the forgetful functor and define Q_0 from $\mathcal{S}^{\mathbf{T}}$ into topological $\mathbf{T}_0$-algebras by using the Q topology together with U_0 for the $\mathbf{T}_0$-structure. By 1.2, 1.3 and 3.15 we see that Q_0 is a full embedding. If $X \in \mathcal{S}^{\mathbf{T}}$ then $Q_0(X)$ is complete, Hausdorff, $\mathfrak{M}$-generated and linearly compact by 1.5 and the above results. Moreover suppose $K \subseteq X$ is a closed $\mathbf{T}_0$-kernel with X/K algebraically equivalent to a model. Then $X \to X/K$ is a **T**-homomorphism by 1.5(c), so $Q(X) \to Q(X/K)$ is continuous. But $Q(X/K)$ is discrete so K is open. Hence X is strongly $\mathfrak{M}$-generated. Also $\mathcal{S}^{\mathbf{T}} = \mathfrak{M}$-objects by 3.16. Conversely let X be a complete, Hausdorff strongly $\mathfrak{M}$-generated $\mathbf{T}_0$-algebra. Let $\{U_\alpha\}$ be the base of all open $\mathbf{T}_0$-kernels for which $X/U_\alpha \in \mathfrak{M}$. Then $\{X/U_\alpha\}$ is a regular filtered, partially ordered diagram in the obvious way. Since X is Hausdorff and complete $X = \text{Top-lim}\{X/U_\alpha\}$ (or more precisely the limit in topological $\mathbf{T}_0$-algebras). We claim that $Q(\mathbf{T}\text{-lim}\{X/U_\alpha\}) = \text{Top-lim}\{X/U_\alpha\}$ which shows $X = Q_0(\mathbf{T}\text{-lim}\{X/U_\alpha\})$, hence we can regard $X \in \mathcal{S}^{\mathbf{T}}$. Let $f : \mathbf{T}\text{-lim}\{X/U_\alpha\} \to M \in \mathfrak{M}$ be a **T**-homomorphism. Then Ker f closed in the limit topology by 3.14 hence is open as X is strongly $\mathfrak{M}$-generated. The claim now follows, from 1.5.

(b) In view of the remarks preceeding the statement of Theorem 3.2 $\widetilde{T}(X)$ is linearly compact for all $X \in$ Top hence every $\widetilde{\mathbf{T}}$-algebra is linearly compact, therefore complete. Since every regular diagram is **T**-regular for $\mathfrak{M}$, we see that much of (b) follows from 2.4 and 2.5. That φ is a coreflection and behaves as stated follows from 1.5 and 3.14 and the observation that φ preserves models and their limits and everything is a limit of models.

(c) Let $A \subseteq X$ be a **T**-subalgebra of the **T**-algebra X. Let $\alpha : A \to X$ be the inclusion and let $x \in X$ be in the closure of $\alpha(A)$. Let H be the set of all **T**-homomorphism $h : A \to M$ which factor through α as $h = g\alpha$ where $g : X \to M$ is onto and **T**-admissible. Then for each $h \in H$ let $\gamma(h) = g(x)$ where g is chosen so that $h = g\alpha$ (note that $g(x)$ is uniquely determined even though g is not). Claim that (H, γ) is finitely consistent. Let $h_1, \ldots, h_n \in H$ be given and chose g_i with $h_i = g_i\alpha$. Then $\cap g_i^{-1}(g_i(x))$ is an open

neighbourhood of x so there exists $a \in A$ with $g_i(\alpha a) = g_i(x)$ or $h_i(a) = \gamma(h_i)$ for $i = 1, \ldots, n$. By applying 3.12 to A there exists $a \in A$ with $g(a) = g(x)$ for all $g: X \to M$. Since X is Hausdorff $a = x$ which shows that A is closed. q.e.d

Proof of Theorem 3.3. (a) We have LCC so Theorem 3.2 applies and it clearly suffices to show that every $\mathbf{T}$-algebra is strongly $\mathcal{M}$-generated. Let $X \in \mathrm{Top}^{\mathbf{T}}$ and K a closed $\mathbf{T}_0$-kernel with X/K algebraically equivalent to a model be given. Then there exists a smallest $\mathbf{T}_0$-kernel U of X for which U is open and $K \subseteq U$ by applying the d.c.c. (or the minimum condition) for X/K. By 1.4(c) we have $K = U$ so K is open

(b), (c) and (d). By (a) above and the fact that φ preserves limits and models we see that the $\tilde{U}$-topology on $\lim M_i$ coincides with the Q-topology. Thus Q preserves limits of models hence 2.3 applies. Q preserves all limits as Q has a left adjoint by 3.2(a) and 2.3(c). Also (c) follows straightforwardly from 3.2(c) and 2.3(b), and (d) is the same as 2.3(d).

(e) The idea of the proof is to obtain a "continuous version" of 3.12. Let $f: A \to B$ be a given onto $\mathbf{T}$-homomorphism. Note that by 3.16 we have $A = \hat{T}(A) = \lim\{M_g \mid g \in (A, I)_0\}$ where $g: A \to M_g$ is in $(A, I)_0$ iff g is admissible and onto. We define (H, γ) to be a $\wedge$*-closed continuous-B-prescription* if $H \subseteq (A, I)_0$ and for each $h \in H$, $\gamma(h)$ is a continuous function from B to M_h. Moreover if $b \in B$ we let (H, γ_b) be the ordinary prescription with $\gamma_b(h) = \gamma(h)(b)$. It is further required that (H, γ_b) be $\wedge$-closed for all $b \in B$.

Now let (H, γ) be such a continuous prescription and $g: A \to M_g$ be arbitrary in $(A, I)_0$ with $g \notin H$. Choose $h \in H$ such that $g(\mathrm{Ker}\, h)$ is a minimum in M_g. Let $M_0 = M_g/g(\mathrm{Ker}\, h)$ and let $p: M_g \to M_0$ be the projection. Observe that there exists $r: M_h \to M_0$ with $rh = pg$. It is readily shown, from the choice of h, that $\hat{\gamma}_b(g)$ is a coset of precisely $g(\mathrm{Ker}\, h)$ for each $b \in B$. Moreover, if $m \in \hat{\gamma}_b(g)$ then $p(m) = r(\gamma_b(h))$ (in view of 3.7) so $p^{-1}(r\,\gamma_b(h)) = \hat{\gamma}_b(g)$ as they are both cosets of $g(\mathrm{Ker}\, h)$ and have points in common. (Note $r\,\gamma_b(h)$ is a single point as $\gamma(h)$ is a function.) Thus if $s: M_0 \to M_g$ is any section (that is $ps = M_0$ and s is *not* required to be a $\mathbf{T}$-homomorphism) then (H, γ) can be extended to a $\wedge$-closed continuous extension (G, β) where $\beta(g) = sr\,\gamma(h)$ as the method used in 3.8 can be applied pointwise. By Zorn's Lemma, every $\wedge$-closed continuous B-prescription can be extended to one with $H = (A, I)_0$ which gives rise to a continuous map $\mu: B \to A$ (that is $\mu: Q(B) \to Q(A)$ but is not necessarily $\mathbf{T}$-admissible) such that $h\mu = \gamma(h)$ for all $h \in (A, I)_0$. Now observe that $B = \hat{T}(B) = \lim\{M_k \mid k \in (B, I)_0\}$. Let $H_0 = \{kf \mid k \in (B, I)_0\}$ and define γ_0 so that $\gamma_0(kf) = k$ (which is determined as f is onto).

Clearly (H_0, γ_0) is a $\wedge$-closed continuous B-prescription so there exists a continuous $s: B \to A$ such that $kfs = k$ for all $k \in (B, I)_0$ which implies $fs = B$. Finally the d.c.c. hypothesis is implied by (b) and (c). Let $M \in \mathcal{M}$ and $A_1 \supseteq A_2 \supseteq \ldots$ be a descending chain of $\mathbf{T}_0$-kernels. By (b) and (c) the image of M in $\prod M/A_n$ has the relative topology induced by the product topology. But the image of M must be discrete by 1.5(a) which implies that the descending chain is ultimately constant. Moreover, given LCC it is readily shown that (a) ⇒ (b) ⇒ (d) by the above arguments. Using

3.2 we can readily obtain (d) ⇒ (a) and (b) ⇒ (c). Thus (a), (b) and (d) each individually imply (b) and (c). q.e.d.

§4. Examples

Most of the applications of Theorem 3.1 are immediate (e.g. the equational completion of finite sets is the category of compact Hausdorff spaces and the corresponding $\mathrm{Top}^{\mathbf{T}}$ is the subcategory of totally disconnected members of $\mathcal{S}^{\mathbf{T}}$. Also the equational completion of finite groups is pro-finite groups which is tripleable over Top, cf. [7] and [8]). Applications of Theorems 3.2 and 3.3 sometimes require some preliminary work with individual operations to get a normal separating triple as 4.1 and 4.2 show. A number of interesting papers have been written on linearly were an equational completion led to Theorems 3.2 and 3.3. Example 4.4 shows that even if Q grossly fails to preserve limits the topology can still be very useful in des-
even if Q grossly faisl to preserve limits the topology can still be very useful in des-
cribing the equational completion. In 4.5 the subject is briefly discussed from the point of view of operations.

Notation. Unless otherwise specified, rings are assumed to have units and ring homomorphisms and modules are unitary. In contrast to the previous sections, the group operations here happen to be Abelian and are denoted by "+". Also "0" instead of "1" is used for the group identity, so one must exercise care in applying the previous results.

4.1. Fields. Let $\mathcal{F}_0$ be the category of fields and (unitary) ring homomorphisms. As it stands $\mathcal{F}_0$ does not admit an obvious separating triple (as subrings of fields are not always fields) nor does $\mathcal{F}_0$ have finite products. We first observe that if $\mathcal{F}$ is the category of all finite products of fields (and unitary ring homomorphisms) then $\mathcal{F}$ and $\mathcal{F}_0$ have the same equational completion. (It can readily be shown that if $\prod K_i \in \mathcal{F}$ and if $F \in \mathcal{F}_0$ then any homomorphism $\prod K_i \to F$ factors through a projection $\prod K_i \to K_i \to F$. Hence every operation of $\mathcal{F}_0$ is also an operation of $\mathcal{F}$.)

We next observe that the theory of commutative regular rings is a separating triple for $\mathcal{F}$. In any ring R we say that r is the *semi-inverse* of s if $rsr = r$, $srs = s$ and $sr = rs$. If the *semi-inverse* exists it is determined by the ring structure. (If t is another semi-inverse of r then $tr^2 = r = sr^2$ and $t = t^3r^2 = t^2sr^2 = s^3r^2 = s$.) It follows that every ring homomorphism preserves semi-inverses when they exist, hence the semi-inverse is an operation in any category of rings with semi-inverses. A *commutative* ring has semi-inverses for every element iff it is regular (for each r there exists an r' with $rr'r = r$).

Every field is clearly regular (with 0 its own semi-inverse) and every member of $\mathcal{F}$ is regular. Moreover it can be easily verified that a subregular ring of a finite product of fields is in $\mathcal{F}$. (Look at idenpotents.) Also $\mathcal{F}$ is closed under the formation of

quotients (this can be done by considering idempotents or by observing that the members of $\mathcal{F}$ are precisely the commutative regular Artin rings). Thus the theory of commutative regular rings is a normal separating triple for $\mathcal{F}$.

We now can apply Theorems 3.2 and 3.3 (as every member of $\mathcal{F}$ has d.c.c. for ideals). Thus the equational completion of $\mathcal{F}$ is the category of regular pro-objects. It is also the category of topological limits of members of $\mathcal{F}$, hence also of $\mathcal{F}_0$. Clearly if $X \in \mathcal{S}^T$ then $X = \lim K_i$ and it can be arranged that K_i is a field and the projection map $X \to K_i$ is onto, for all i. It follows that every map in the diagram $\{K_i\}$ is onto hence is an isomorphism. Thus X is a product of fields. Thus the equational completion of the category of fields is the category of all products of fields and ring homomorphisms which are continuous in the product topology.

(The same argument works if one allows non-unitary homomorphisms and/or skew-fields or ordered fields (which have a lattice operation).)

The category of fields obviously has a large number of weird operations. For example one can define $w(x, y) = x + y$ in characteristic 0 fields but $w(x, y) = xy$ elsewhere. In fact if we wish to describe an arbitrary n-ary operation, w, we have to consider all collections of n-tuples $(x_1, \ldots, x_n)$ in fields F. One obvious restriction on $w(x_1, \ldots, x_n)$ is that it must lie in the subfield generated by $x_1, \ldots, x_n$ and this is essentially the only restriction on w. Formally, define an *n-pointed field* as a field F together with an n-tuple $(x_1, \ldots, x_n)$ such that no proper subfield of F contains each x_i. Let $\mathcal{F}(n)$ be a representative collection of n-pointed fields such that each n-pointed field is isomorphic to a unique member of $\mathcal{F}(n)$ (where isomorphisms must preserve the n-tuple). Then one can define an operation w by choosing $w(x_1, \ldots, x_n) \in F$ entirely at random for each $[F, (x_1, \ldots, x_n)] \in \mathcal{F}(n)$. (Notice that F has no non-trivial automorphism preserving the n-tuple.) In other words $T(n) = \prod \mathcal{F}(n)$ (despite the notation we have not assumed that n is finite). It clearly follows that the theory of fields does not have rank.

It would be of some interest perhaps to describe the theory generated by finitary field operations. Consider for example the unary operation r such that $r(x) = 1$ iff x is rational (in the usual characteristic 0 sense) and $r(x) = 0$ otherwise. (This defines r on fields and by an obvious extension r is defined for all products of fields.) Not every ring homomorphism $f: Q^N \to K$ preserves r (where Q^N is a countable product of rationals and K is any field). In fact f is continuous iff f preserves r (in this case). This refutes the conjecture that the topology is used only to handle infinitary operations. Also the theory generated by the finitary operations is richer than the theory of commutative regular rings.

4.2. Artin rings. In [5] a *pseudocompact ring* is defined as a complete Hausdorff ring R which admits a base at 0 of two-sided open ideals I for which R/I is an Artin ring. It is easily seen that in our terminology this is the same thing as an "Artin-generated" Hausdorff, linearly compact ring or, alternatively a regular pro-object for Artin rings.

In view of Theorems 3.2 and 3.3 it seems reasonable to conjecture that the cate-

gory of pseudocompact rings is the equational completion of the category of Artin rings. We shall show that this is not quite the case but that for certain classes $\mathscr{A}$ of Artin rings the $\mathscr{A}$-generated pseudocompact rings is the equational completion of $\mathscr{A}$

Notice that the Artin rings are closed under finite products but not under the formation of subrings so the theory of rings is not a separating triple. Moreover as implied by the example below there is no way of enriching the theory of rings so as to obtain a separating triple. There are however separating triples for certain classes of Artin rings.

If the category of all pseudocompact rings were equational the underlying set functor would preserve and create limits hence the discrete pseudocompact rings (i.e. the Artin rings) would have to be closed under finite limits and arbitrary intersections. The following counter-example shows this is not so. Let A be the Artin ring of all 2 by 2 real matrices. For convenience, for the remainder of the paragraph, let $\langle r, s\rangle$ denote the 2 by 2 matrix $(a_{ij}) \in A$ for which $a_{11} = a_{22} = r$ and $a_{21} = 0$ and $a_{12} = s$. Let $P = \langle 1, 1\rangle$. Define $\Phi(X) = PXP^{-1}$ then $\Phi: A \to A$ is a ring homomorphism. Let E be the equalizer of Φ and the identity on A. Then E is the ring of all matrices of the form $\langle a, b\rangle$. Let C be the complex numbers and let $\theta: E \to C$ be defined by $\theta\langle a, b\rangle = a$. Then the intersection of all equalizers of θ and $\gamma\theta$ (for γ an automorphism of C) is the ring E' of all $\langle a, b\rangle$ for which a is rational and b is real. Then E' is not an Artin ring. (Observe that the reals are an infinite dimensional vector space of the rationals. If V is any linear subspace, then the set of all $\langle 0, b\rangle$ with $b \in V$ is an ideal of E'. Clearly the ideals of E' do not satisfy the d.c.c. nor even the a.c.c..)

We now construct a triple $\mathbf{T}_0$ whose algebras shall be rings R endowed with a 3-tuple of unary operations that map x into (x_1, x_2, x_3) such that

$$x = x_1 + x_2$$

$$x_1 x_2 = x_2 x_1 = 0$$

x_3 is the semi-inverse of x_1.
(that is $x_1 x_3 x_1 = x_1$, $x_3 x_1 x_3 = x_3$ and $x_1 x_3 = x_3 x_1$).

A $\mathbf{T}_0$-algebra R shall be called *canonical* if x_2 is nilpotent for all $x \in R$. For a canonical $\mathbf{T}_0$-algebra the operations are uniquely determined by the ring structure (and the requirement that x_2 be nilpotent). Thus the $\mathbf{T}_0$-homomorphisms between canonical $\mathbf{T}_0$-algebras are precisely the ring homomorphisms. (To prove this, let R be a canonical $\mathbf{T}_0$-algebra and assume $x = r+p$ where p is nilpotent, r has semi-inverse s and $rp = pr = 0$. Claim $r = x_1$, $p = x_2$ and $s = x_3$. Observe that for sufficiently large m we have $x^m = r^m = x_1{}^m$. This implies $s^m = x_3{}^m$. Hence $x_1 = x_1{}^{m+1} x_3{}^m = r^{m+1} s^m = r$. Therefore, $s = x_3$ and $p = x_2$.)

A ring shall be called *canonical* if it can be endowed with a canonical $\mathbf{T}_0$-structure. Notice that canonical rings are closed under finite products, quotients, the formation of $\mathbf{T}_0$-subalgebras (but *not* under infinite products). Moreover every simple

Artin ring (i.e. ring of all linear transformations of a finite dimensional vector space) is canonical in view of Fitting's Lemma.

We now construct a class $\mathscr{A}$ of Artin rings. Start with all canonical rings which contain a minimal field (i.e. either the rationals or Z_p for some prime p) as a (unitary) subring and which are finite dimensional as vector spaces over the minimal field. To this collection of rings adjoin all finite products to get $\mathscr{A}$. Then every member of $\mathscr{A}$ is an Artin ring and $\mathbf{T}_0$ is a separating triple for $\mathscr{A}$. Thus the $\mathscr{A}$-generated pseudocompact rings is the equational completion of $\mathscr{A}$ and is also the category of pro-object for $\mathscr{A}$ by 3.3 and 2.2. (Each object of $\mathscr{A}$ satisfies d.c.c. for both $\mathbf{T}_0$-subalgebras and $\mathbf{T}_0$-kernels.)

One can also adjoin all fields and skew fields to $\mathscr{A}$ and close up under finite products. Then the resulting rings are still Artin and $\mathbf{T}_0$- is still a separating triple (so 3.3 still applies) but d.c.c. for $\mathbf{T}_0$-subalgebras is lost. If any additional simple Artin rings are adjoined then the argument of the above counter-example will apply.

4.3. Linearly compact modules. Let R be any ring. Then the (discrete) linearly compact R-modules have an equational completion that can be computed by 3.2. The theory of R-modules is still a separating triple even if one treats the more general case of topological R-modules for a topological ring R (as done in [18]) provided the models are always discrete, topological linearly compact R-modules. A number of interesting properties of linearly compact modules can be found in [13] and [18]. The case in which R is a pseudocompact ring is treated in [5]. Then a discrete topological R-module has the d.c.c. iff it is of finite length. The case of pseudocompact algebras is also discussed in [5] and provides another straightforward example for Theorem 3.3.

4.4. Countable sets (a realcompact case). Let $\mathscr{M}$ be the category of countable sets and all functions. Let $I: \mathscr{M} \to \mathscr{S}$ be the inclusion. Then the theory of sets is a separating triple. Moreover, as we shall point out, the Q topologies are all Hausdorff, hence the equational completion is a full subcategory of Top. The main step in computing this full subcategory are:

Definition. A topological space is *nearly discrete* if it is Hausdorff and if every non-measurable collection of open sets has open intersection.

Lemma. (a) *A nearly discrete space is realcompact iff every open cover has a non-measurable subcover.*

(b) *The nearly discrete spaces are coreflective in the category of Hausdorff spaces. The nearly discrete, realcompact spaces are closed under the formation of (coreflected) limits.*

(c) *The category of nearly discrete, realcompact spaces is tripleable over $\mathscr{S}$.*

Proof (sketch). Most of the proof follows by mimicking the corresponding proofs for compact Hausdorff spaces. The coreflection into nearly discrete spaces is accomplished by letting the class of all non-measurable intersections of open sets be a base for the new topology. If X is realcompact then it is a closed subset of R^n for some n. Moreover X is still closed in the nearly discrete topology on R^n so X has the covering property (which is inherited by nearly discrete products and closed subsets). Conversely the covering property implies realcompactness using the z-ultrafilter characterization of realcompact spaces (see ref. [9]). q.e.d.

Corollary. *The equational completion of the category of countable sets is the category of realcompact, nearly discrete topological spaces. (If every cardinal is non-measurable this is the category of discrete spaces or simply S in effect.)*

Proof (sketch). If X is realcompact then X can be embedded as a closed subset of a product of real lines [9]. If, in addition, X is nearly discrete, then it easily follows that X is equivalent to a closed subset of a (coreflected) product of discrete countable spaces. Using the fact that the continuous image of a realcompact space is (in the category of nearly discrete spaces) realcompact, hence closed, X is the canonical limit of discrete countable spaces. The result now follows straightforwardly. q.e.d.

4.5. What do the operations look like? We have dealt with the general theory of equational completions primarily from the point of view of triples. At this point we collect a few observations about the operations. Recall that the set of n-ary operations is $T(n)$, hence we have a topological space of n-ary operations. Thus the corresponding equational theory $\mathcal{T}$ (say, the category with $\mathcal{T}(n, k) = T(n)^k$) can be regarded as a category with topological hom sets. However, it is *not* a topological category in the sense of Beck [3] since composition $T(n)^k \times T(k) \to T(n)$ although continuous for all finite k (which can be shown by using the type of argument given in 1.2) is nonetheless almost never continuous for infinite k. However, for any fixed k-tuple on a **T**-algebra X the evaluation map $T(k) \to X$ is continuous (as it is **T**-admissible).

We know that $T(n) = \lim(n, I)$. Moreover if $\mathbf{T}_0$ is any normal separating triple and if LCC is satisfied then $T(n) = \lim(n, I)_0$ and for every full, filtered subdiagram $\Delta \subseteq (n, I)_0$ the canonical map $T(n) \to \lim \Delta$ is onto (using the argument of 3.10). Thus one can obtain an n-ary operation by prescribing its behaviour on a set Δ of n-tuples and lifting it from $\lim \Delta$ to $T(n)$ (the lifting depends on a choice). Moreover if the models all satisfy d.c.c. for $\mathbf{T}_0$-kernels then by (b) and (e) of Theorem 3.3, the entire space, $\lim \Delta$, can be *continuously* lifted to a topologically equivalent closed subspace of $T(n)$.

Our final observation is that the equational completions generally do not have rank. For example suppose that the hypotheses of 3.3 are satisfied and that **T** has (infinite) rank r. Let $X \subseteq Y$ be any **T**-subalgebra and let s be the (cardinal) successor of r. Let $A \subseteq Y^s$ be the set of all points which project into X for all but at most r

projections. Then A would be a non-closed $\mathbf{T}$-subalgebra contradicting 3.2(c). The same contradiction works for the compact case. It also works for countable sets iff at least one measurable cardinal exists. We note that here the n-ary operations are the ω-ultrafilters on n.

Note. The discontinuity of the property infinitary operations has led us to restrict ourselves at times to the case where the separating triple is finitary. However, all hypotheses of the form "$\mathbf{T}_0$ is finitary" can be eliminated if we agree that a topological $\mathbf{T}_0$-algebra is to be a $\mathbf{T}_0$-algebra with topology such that the *finitary* operations are continuous.

References

[1] H. Appelgate and M. Tierney, Categories with models, Lecture Notes in Mathematics, Vol. 80 (Springer-Verlag, 1969) 156–244.

[2] M. Artin and B. Mazur, Etale Homotopy II, Lecture Notes in Mathematics, Vol. 100 (Springer-Verlag, 1969).

[3] J. Beck, On H-spaces and infinite loop spaces, Lecture Notes in Mathematics, Vol. 99 (Springer-Verlag, 1969) 139–153.

[4] J. Beck, untitled manuscript.

[5] A. Brumer, Pseudocompact algebras, profinite groups and class formations, J. Algebra 4 (1966) 442–470.

[6] P. Freyd, Abelian categories (Harper and Row, 1964).

[7] D. Gildenhuys, On varieties of groups and inverse limit groups.

[8] D. Gildenhuys and Chong-Keang Lim, Free pro-$\mathcal{C}$ groups, Math. Zeitschrift, to appear.

[9] L. Gilman and M. Jerison, Rings of continuous functions (D. Van Nostrand, Princeton, 1960).

[10] L. Henkin, A problem on inverse mapping system, Proc. Am. Math. Soc. 1 (1950) 224–225.

[11] J. Lambek, Lectures on modules and rings (Blaisdell, Waltham, 1966).

[12] S. Lefschetz, Algebraic topology, Colloquim Publications, Vol. 27 (Am. Math. Soc. New York, 1942).

[13] H. Leptin, Linear kompakte Moduln and Ringe, Math. Zeitschrift 62 (1955) 241–267.

[14] F. Linton, An outline of functorial semantics, Lecture Notes in Mathematics, Vol. 80 (1960) 7–52.

[15] F. Linton, Some aspects of equational categories, Proc. Conf. on Categorial algebra, LaJolla 1965 (Springer-Verlag, 1966) 84–94.

[16] F. Manes, A triple theoretic construction of the compact algebras, Lecture Notes in Mathematics, Vol. 80 (Springer-Verlag, 1969) 91–118.

[17] J.-P. Serre, Cohomologie Galoisienne, Lecture Notes in Mathematics, Vol. 5 (Springer-Verlag, 1965).

[18] D. Zelinsky, Linearly compact modules and rings, Amer. J. Math. 75 (1953) 79–90.

JOURNAL OF PURE AND APPLIED ALGEBRA Volume 1, No. 4 (1971) 347–360

DESCENDING CHAINS AND THE KRULL ORDINAL OF COMMUTATIVE NOETHERIAN RINGS

Hyman BASS
Dept. of Mathematics, Columbia University
New York, N.Y. 10027, U.S.A.

Received 6 May 9171

Introduction

Let A be a commutative ring. It is Noetherian if every totally ordered set C of non zero ideals of A is well ordered by reverse inclusion. In this case C has an ordinal type, ord C. We denote by $o(A)$ the supremum of ord C as C varies over all descending chains of non zero ideals of A.

Eklof and Sabbagh [2, Remark 12] have asked whether there exists a countable ordinal α such that $o(A) < \alpha$ for all commutative Noetherian rings A. They observe that the existence of such an α would result if the class of commutative Noetherian rings could be defined by a sentence in the logical language $L_{\omega_1\omega}$ (see [2, §1.0] for this notation).

Let $\kappa = \kappa(A)$ denote the Krull dimension of A. If κ is finite then one has (Theorem 2.12): $\omega^\kappa \leq o(A)$, with equality if A is an integral domain, and $\omega^{\kappa+1} = o(A)\cdot\omega$. This result was proved independently by Schneider [6]. It shows that $\alpha = \omega^\omega$ suffices above for rings of finite Krull dimension.

When κ is no longer finite we are led to interprete the Krull dimension $\kappa(A)$ as an ordinal, as in Krause [3] (see 2.7) and then the above result (Theorem 2.12) remains valid provided $\kappa(A)$ is *countable*. Thus the question raised by Eklof and Sabbagh is equivalent to the following: Does there exist a countable ordinal α such that every commutative Noetherian ring has Krull ordinal $< \alpha$? We conjecture (in 2.9) that this is so, possibly with $\alpha = \omega\cdot\omega$, or even $\omega + \omega$. The main difficulty here is the scarcity of examples of infinite dimensional Noetherian rings. Simple variants of Krull's original example [4, Staz 19] seem to be the only ones known.

We are able at least to show that a totally ordered set of ideals in any commutative Noetherian ring is countable (Theorem 1.1). Nagata has indicated in a communication to Eklof and Sabbagh that he also has proved this result, but I have not seen his proof.

In the course of these considerations a combinatorial problem on ordinals arose. Its solution, which occupies §3, was worked out jointly with Henryk Hecht (Colum-

bia University). I am grateful to him for permission to reproduce the details here. The problem can be heuristically stated as follows: Let α and β be ordinals. Let A and B be stacked decks of cards of order types α and β, respectively. Let C be the deck obtained by some shuffling together of A and B. How large can one make the ordinal type of C by a suitable shuffling?

§1. Noetherian chains are countable

1.1. This section is devoted to the proof of the following result.

Theorem. *Let A be a commutative Noetherian ring and let M be a finitely generated A-module. Then every set of submodules of M which is totally ordered by inclusion is countable.*

The proof is preceded by a few preliminaries on partially ordered sets.

1.2. Partially ordered sets. Let S and T be partially ordered sets. We shall write

$$S \leq T$$

if there is an order preserving function $f: S \to T$ such that the restriction of f to any totally ordered subset of S is injective.

We denote by

$$S < + T$$

the disjoint union of S and T ordered so as to induce the given orderings in S and in T and so that each element of S precedes each element of T. This operation is associative but *not commutative*.

We denote by

$$S \Delta T$$

the cartesian product $S \times T$ ordered so that $[(s, t) \leq (s', t')] \Leftrightarrow [s \leq s' \text{ and } t \leq t']$. This operation is associative and commutative (up to canonical isomorphism).

1.3. Lemma. *Let S and T be partially ordered sets in which every totally ordered subset is countable. Then the same is true of $S \Delta T$.*

(The converse is true, and trivial, if S and T are non empty. Similar statements for $S < + T$ in place of $S \Delta T$ are likewise trivally verified.)

To prove the lemma consider a totally ordered subset U of $S \Delta T$. The projection U_S (resp. U_T) of U in S (resp. T) is totally ordered, and hence countable, by hypothesis. Since $U \subset U_S \times U_T$ the set U is countable

1.4 The partially ordered set of submodules. Let A be a ring. For any (say left) A-module M denote by

$$S(M)$$

the set of submodules of M ordered by *reverse inclusion*, i.e. $N_1 \le N_2 \Leftrightarrow N_2 \subset N_1$.

Lemma. *If* $0 \to M' \xrightarrow{i} M \xrightarrow{p} M'' \to 0$ *is an exact sequence of A-modules we have*

$$S(M'') \mathbin{<\!+} S(M') \le S(M) \le S(M'') \triangle S(M').$$

Define $f: S(M'') \mathbin{<\!+} S(M') \to S(M)$ by $f(N) = i(N)$ if $N \in S(M')$ and $f(N) = p^{-1}(N)$ if $N \in S(M'')$. Evidently f is an order preserving injection, whence the first inequality.

Define $g: S(M) \to S(M'') \triangle S(M')$ by $g(N) = (p(N), i^{-1}(N))$. Then g is order preserving clearly. The fact that the restriction of g to a totally ordered subset of $S(M)$ is injective follows easily from the 5-lemma. Thus the lemma is proved.

1.5. Proof of Theorem 1.1. Let A and M be as in the theorem. Using a composition series of M as in [1, §1, no. 4, Th. 1] one concludes easily, with the aid of Lemmas 1.3 and 1.4, that it suffices to treat the case when $M = A$ and A is an integral domain.

By Noetherian induction we may further assume that the theorem holds for all proper quotients of A.

Let C be a totally ordered set of non zero ideals in A. Since A is Noetherian C is a well ordered descending chain, and we must show that C is countable. Assume the contrary.

If I is a non zero ideal of A then $\{J \in C \mid I \subset J\}$ is countable since it corresponds to a totally ordered set of ideals of A/I, and the theorem is assumed valid for A/I. It follows that

(1) $$\bigcap_{I \in C} I = 0$$

and that:

(2) There is no sequence $I_1, I_2, I_3, \dots$ of elements of C such that $\bigcap_{n \ge 1} I_n = 0$.

We now define the set $M(C)$ of "minimal prime ideals associated with C" to be the set of minimal elements of $\bigcup_{I \in C} \mathrm{Ass}(A/I)$. Since the prime ideals of A satisfy the descending chain condition [4, Ch. I, Th. 9.3] $M(C)$ is not empty. Suppose $\mathfrak{p} \in M(C)$. There exists an $I \in C$ such that $\mathfrak{p} \in \mathrm{Ass}(A/I)$. If $J \in C$ and $J \subset I$ then there is a $\mathfrak{q} \in \mathrm{Ass}(A/J)$ such that $\mathfrak{q} \subset \mathfrak{p}$ [1, §1, no. 3, Cor. 1 of Prop. 7], and so $\mathfrak{q} = \mathfrak{p}$ by the minimality of $\mathfrak{p}$. Thus we see that $\mathfrak{p} \in M(C) \Leftrightarrow \mathfrak{p}$ is a minimal element of $\mathrm{Ass}(A/I)$ for all sufficiently small ideals $I \in C$.

We contend

(3) $M(C)$ is finite.

If not let $\mathfrak{p}_1, \mathfrak{p}_2, \ldots$ be a sequence of distinct elements of $M(C)$ and choose $I_n \in C$ so that $\mathfrak{p}_n \in \mathrm{Ass}(A/I_n)$. By (2) we have $\bigcap_n I_n \neq 0$ so, by (1), there is a $J \in C$ such that $J \subset I_n$ for all n. But then, as the last paragraph shows, $\mathfrak{p}_n \in \mathrm{Ass}(A/J)$ for all n. This is impossible since $\mathrm{Ass}(A/J)$ is finite, so (3) is established.

For each ideal $I \in C$ we can write $I = I' \cap I''$ where I' is the intersection of the primary components of I associated to prime ideals in $M(C)$, and where each prime in $\mathrm{Ass}(A/I'')$ strictly contains an element of $M(C)$. This is easily seen from a primary decomposition of I. We can further arrange by (transfinite) induction that the ideals I'' (which are not uniquely determined by I in general) also form a totally ordered set; if not one successively replaces I'' by $I'' \cap (\bigcap_J J'')$, where J varies over ideals in C strictly containing I.

We claim:

(4) The chain of ideals $(I')_{I \in C}$ is stationary for sufficiently small $I \in C$

Put $S = A - (\bigcup_{\mathfrak{p} \in M(C)} \mathfrak{p})$ and $B = A[S^{-1}]$. Then B is a semilocal ring. For all $I \in C$ we have $IB = I'B$ and $IB \cap A = I'$ [1, §2, no. 4]. Moreover B/IB has finite length, so it follows that each term of the descending chain $(IB)_{I \in C}$ of ideals of B has only finitely many precedessors. Thus we can find a sequence (I_n) in C such that the sequence $(I_n B)$ is cofinal in $(IB)_{I \in C}$. Since $I' = IB \cap A$ for all $I \in C$ it follows that the sequence (I'_n) is cofinal in the chain $(I')_{I \in C}$. By (2) we have $0 \neq \bigcap_n I_n \subset \bigcap_n I'_n$. It follows that I' is independent of I for all $I \in C$ such that $I \subset I_n$ for all n. Such ideals I exist by (1), so (4) is proved.

Let J denote the (constant) value of I' for all sufficiently small $I \in C$ (see (4)) and let $C_1 = \{I \in C \mid I' = J\}$. Since C_1 is an end segment of the chain C it is still uncountable. Put $D(C) = \{I'' \mid I \in C_1\}$. Then we can write $C_1 =$ "$J \cap D(C)$" $=$ $\{J \cap I'' \mid I'' \in D(C)\}$. The chain $D(C)$ is uncountable, and each element of $\bigcup_{I'' \in D(C)} \mathrm{Ass}(A/I'')$ strictly contains an element of $M(C)$. It follows that each element of $M(D(C))$ strictly contains one of $M(C)$.

Since the chain of ideals $D(C)$ satisfies the same conditions assumed for C we can iterate the above constructions to obtain a sequence of chains

$$C, D(C), D^2(C), \ldots$$

and of non empty finite sets of prime ideals

$$M(C), M(D(C)), M(D^2(C)), \ldots .$$

Further the construction shows that each element of $M(D^{n+1}(C))$ strictly contains one of $M(D^n(C))$.

Put $M = \bigcup_n M(D^n(C))$. Since M is infinite and each element of M contains one of the finite set $M(C)$ there is a $\mathfrak{p}_0 \in M(C)$ such that $M_1 = \{\mathfrak{p} \in M \mid \mathfrak{p}_0 \subsetneqq \mathfrak{p}\}$ is infinite. Each element of M_1 contains one of the finite set $M(D(C)) \cap M_1$ so there is a $\mathfrak{p}_1 \in M(D(C)) \cap M_1$ such that $M_2 = \{\mathfrak{p} \in M_1 \mid \mathfrak{p}_1 \subsetneqq \mathfrak{p}\}$ is infinite. Continuing in

this fashion one constructs a sequence of ideals $\mathbf{p}_n \in M(D^n(C))$ such that $\mathbf{p}_0 \subsetneqq \mathbf{p}_1 \subsetneqq \mathbf{p}_2, \ldots$. This contradicts the fact that A is Noetherian, so Theorem 1.1 is proved.

§2. The Krull ordinal $\kappa(A)$ and its relation to o(A) and to O(A)

For a commutative Noetherian ring A the supremum of the ordinal types of descending chains of non zero ideals is denoted $o(A)$ (see 2.10). We also introduce the Krull ordinal $\kappa(A)$ (see 2.7); when finite it is the usual Krull dimension. The main result of this section (Theorem 2.12) asserts, under the assumption that $\kappa(A)$ is countable, that $\omega^{\kappa(A)} \leqslant o(A)$, with equality if A is an integral domain, and that $o(A) \cdot \omega = \omega^{\kappa(A)+1}$.

It is very likely that the countability assumption on $\kappa(A)$ is automatic (see 2.9 for a more precise conjecture).

The section begins by recalling some notation and elementary facts on ordinals. A complete reference is Sierpinsky [7].

2.1. Ordinals. They are the isomorphism classes, or types, of well ordered sets. The ordinals themselves are well ordered, and the first few are denoted:

$$0, 1, \ldots, n, \ldots, \omega, \omega + 1, \ldots, \omega + \omega, \ldots, \Omega, \ldots$$

If α is any ordinal the set $[\alpha)$ of ordinals $\beta < \alpha$ is a well ordered set of type α. The ordinal ω is the first infinite one, i.e. Card $[\omega)$ is infinite; similarly Ω is the first uncountable ordinal.

2.2. The ordinal of a partially ordered set S. It is the supremum, denoted ord(S), of the types of all well ordered subsets of S. For example $\alpha = \operatorname{ord}[\alpha)$ for any ordinal α.

2.3. Sums. If α and β are ordinals then $\alpha + \beta = \operatorname{ord}([\alpha) <+ [\beta))$ (see 1.2 for the notation $<+$). If S and T are partially ordered sets it is easily seen that

$$\operatorname{ord}(S <+ T) = \sup_U (\operatorname{ord} U + \operatorname{ord} T) \leqslant \operatorname{ord} S + \operatorname{ord} T,$$

where U varies over the well ordered subsets of S. The inequality may be strict.

2.4. Products. If α and β are ordinals the products $\alpha \cdot \beta$ is defined to be ord $([\alpha) \times [\beta))$, where $[\alpha) \times [\beta)$ is ordered (anti)-lexicographically: $[(a, b) < (a', b')] \Leftrightarrow [b < b'$ or $b = b'$ and $a < a']$. Note that it is associative but not commutative. For example $2 \cdot \omega = \omega \neq \omega \cdot 2 = \omega + \omega$.

2.5. Exponentials. Let α and β be ordinals. Denote by $[\alpha)^{([\beta))}$ the set of functions $f: [\beta) \to [\alpha)$ with "finite support" (i.e. $f(b) = 0$ for all but finitely many $b \in [\beta)$). If g is another such function and $g \neq f$ then there is a greatest $b \in [\beta)$ such that $g(b) \neq f(b)$. We write $f < g \Leftrightarrow f(b) < g(b)$. This defines a well ordering of $[\alpha)^{([\beta))}$ (by "last difference"), and one puts

$$\alpha^\beta = \operatorname{ord}([\alpha)^{([\beta))}).$$

For example $\alpha^2 = \alpha \cdot \alpha$. More generally

$$\alpha^{(\beta+\gamma)} = (\alpha^\beta)(\alpha^\gamma)$$

for all ordinals α, β, γ.

2.6. The Krull ordinal of a partially ordered set. Let S be a non empty partially ordered set. We inductively define a subset $D_\alpha(S)$ for each ordinal α, as follows: $D_0(S)$ is the set of maximal elements of S. If $\alpha > 0$ we put

$$D_\alpha(S) = \{p \in S \mid p < q \Rightarrow q \in \bigcup_{\beta < \alpha} D_\beta(S)\}.$$

These sets form an ascending chain, so there is a least ordinal α such that $D_{\alpha+1}(S) = D_\alpha(S)$. If $D_\alpha(S) = S$ we call α the Krull ordinal of S and denote it $\kappa(S)$.

If S is Noetherian then $\kappa(S)$ is defined, i.e. $D_\alpha(S) = S$. For if not let p be a maximal element of $S - D_\alpha(S)$. If $q > p$ then $q \in D_\alpha(S)$ by maximality of p, and so $p \in D_{\alpha+1}(S) = D_\alpha(S)$: contradiction.

Even if S is both Noetherian and Artinian there is no universal bound on the possible size of $\kappa(S)$. For suppose Λ is a set of ordinals and that for each $\alpha \in \Lambda$, S_α is a Noetherian and Artinian partially ordered set such that $\kappa(S_\alpha) = \alpha$. Let S denote the disjoint union of the S_α plus one extra element 0, ordered as follows: Each S_α has its given ordering, elements of S_α are incomparable with those of S_β for $\beta \neq \alpha$, and 0 is the least element of S. It is easily verified that S is Noetherian and Artinian and that $\kappa(S) = \sup \Lambda$ if Λ has no greatest element, and $\kappa(S) = \alpha + 1$ if α is the greatest element of Λ.

2.7. The Krull ordinal of a Noetherian ring. Let A be a non zero commutative Noetherian ring. As with Krause [3] (who considers also non commutative rings) we define the Krull ordinal, denoted $\kappa(A)$, of A to be the Krull ordinal of the set $\operatorname{spec}(A)$ of prime ideals of A, ordered by inclusion. Note that the partially ordered set $\operatorname{spec}(A)$ is both Noetherian and Artinian.

When $\kappa(A)$ is finite it is easily seen to coincide with the usual Krull dimension. In general the Krull ordinal is characterized by the following properties:

0) $\kappa(A) = 0$ if A is a field.

1) $\kappa(A) = \sup \kappa(A/\mathbf{p})$, where $\mathbf{p}$ varies over $\operatorname{spec}(A)$.

2) Suppose A is an integral domain. Put $\alpha = \sup\{\kappa(A/\mathbf{p}) | \mathbf{p} \in \operatorname{spec}(A), \mathbf{p} \neq 0\}$. Then $\kappa(A) = \alpha + 1$ if there is a $\mathbf{p} \neq 0$ in $\operatorname{spec}(A)$ such that $\kappa(A/\mathbf{p}) = \alpha$. Otherwise $\kappa(A) = \alpha$.

The supremum in 1) is actually a maximum since it suffices to consider only the finite set of minimal primes $\mathbf{p}$ of A. In case A is an integral domain it is easily seen that $\kappa(A/I) < \kappa(A)$ for all proper ideals $I \neq 0$ of A.

It appears that the only known examples of commutative Noetherian rings of infinite Krull dimension are simple variants of the original example of Krull [4, Satz 19] reproduced in Nagata [5, Appendix A1, Ex. 1]. This example has Krull ordinal ω because its proper quotients have Krull dimensions which are finite, but unbounded. In view of the next proposition one also has examples of Krull ordinal $\omega + n$ for all integers $n \geq 0$.

2.8. Proposition. *Let A be a commutative Noetherian ring and let $A[X]$ be a polynomial ring in one variable over A. Then $\kappa(A[X]) = \kappa(A) + 1$.*

One reduces the proposition easily to a consideration of A modulo each of its minimal prime ideals, so we may assume A is an integral domain. Put $\kappa = \kappa(A) = \kappa(A[X]/XA[X])$. Since 0 and $XA[X]$ are prime ideals of $A[X]$ it follows that $\kappa(A[X]) \geq \kappa + 1$.

To prove the reverse inequality we must show that $\kappa(A[X]/\mathbf{p}) \leq \kappa$ for all $\mathbf{p} \neq 0$ in $\operatorname{spec}(A[X])$. If $\mathbf{p}_0 = \mathbf{p} \cap A \neq 0$ then $\kappa(A[X]/\mathbf{p}) \leq \kappa((A/\mathbf{p}_0)[X]) = \kappa(A/\mathbf{p}_0) + 1$ (by Noetherian induction) $\leq \kappa$ because $\kappa(A/\mathbf{p}_0) < \kappa$.

Suppose now that $\mathbf{p} \cap A = 0$. It suffices to show that if $\mathbf{p} \subsetneq \mathbf{q}$, $\mathbf{q} \in \operatorname{spec}(A[X])$, then $\kappa(A[X]/\mathbf{q}) < \kappa$. The chain $0 \subsetneq \mathbf{p} \subsetneq \mathbf{q}$ cannot remain a strictly increasing chain of prime ideals after localizing to the field of fractions F of A (for $F[X]$ is a principal ideal domain) so we must have $\mathbf{q}_0 = \mathbf{q} \cap A$ different from zero. Put $J = \mathbf{q}_0 A[X] + \mathbf{p}$. Then $A[X]/\mathbf{q}$ is a quotient of $A[X]/J \cong (A/\mathbf{q}_0)[X]/J'$, where J' is the image of $\mathbf{p}$ modulo $\mathbf{q}_0 A[X]$. If $J' \neq 0$ then $\kappa(A[X]/\mathbf{q}) < \kappa((A/\mathbf{q}_0)[X]) = \kappa(A/\mathbf{q}_0) + 1$ (by Noetherian induction) $\leq \kappa$. Suppose, finally, that $J' = 0$, i.e. that $\mathbf{p} \subset \mathbf{q}_0 A[X]$. This inclusion is strict since $\mathbf{p} \cap A = 0$, so $\operatorname{ht}(\mathbf{q}_0 A[X]) \geq 2$. On the other hand $\operatorname{ht}(\mathbf{q}_0) = \operatorname{ht}(\mathbf{q}_0 A[X])$. For if the obvious inequality, $\operatorname{ht}(\mathbf{q}_0) \leq \operatorname{ht}(\mathbf{q}_0 A[X])$ were strict it would follow that $\kappa(A_{\mathbf{q}_0}[X]) > \kappa(A_{\mathbf{q}_0}) + 1$, contradicting the finite dimensional case of the Proposition (cf. [5, Ch. 1, (9.10)]). Therefore $\kappa(A/\mathbf{q}_0) + 2 \leq \kappa$ and so $\kappa(A[X]/\mathbf{q}) \leq \kappa((A/\mathbf{q}_0)[X]) = \kappa(A/\mathbf{q}_0) + 1 < \kappa$. This concludes the proof.

2.9. Conjecture. *There exists a countable ordinal α such that $\kappa(A) \leq \alpha$ for all commutative Noetherian rings A.* I suspect even that $\alpha = \omega \cdot \omega$ will suffice. $\alpha = \omega + \omega$ even suffices for all the known examples. My evidence for this conjecture is rather weak. It consists mainly of the difficulty I have found in manufacturing new examples of infinite dimensional Noetherian rings. On the other hand I don't even know how to prove that $\kappa(A)$ must be countable.

2.10. The ordinals $o(M)$ and $O(A)$. Let A be a commutative Noetherian ring. For each finitely generated A-module M we denote by $S(M)$ the set of submodules of M, ordered by reverse inclusion (cf. 1.4) and we put

$$o(M) = \operatorname{ord}(S(M) - \{0\})$$

(cf. 2.2). Thus $o(M)$ is the supremum of the ordinal types of descending chains of non zero submodules of M. We further put

$$O(A) = \sup_M o(M)$$

where M varies over all finitely generated A-modules.

2.11. Lemma. $O(A) \leq o(A) \cdot \omega$

Let M be a finitely generated A-module. It suffices to show that $o(M) \leq o(A) \cdot m$ for some integer m. Since M is a quotient of some free A-module A^n it suffices to treat the case $M = A^n$. This we do by induction on n, the case $n = 1$ being immediate. If $n > 1$ we use the exact sequence $0 \to A^{n-1} \to A^n \to A \to 0$. It suffices to know that if $0 \to M' \to M \to M'' \to 0$ is an exact sequence of finitely generated A-modules then $o(M)$ is bounded above by some finite sum of terms equal to $o(M'')$ and $o(M')$. According to 1.4 we have the relation $S(M) \leq S(M'') \triangle S(M')$ on partially ordered sets. Therefore the required inequality follows from 3.7 in § 3 below, which asserts that $\operatorname{ord}(S \triangle T) \leq \sup(\operatorname{ord}(S), \operatorname{ord}(T)) \cdot 3$ for any partially ordered sets S and T.

2.12. Theorem. *Let A be a commutative Noetherian ring. Assume the Krull ordinal $\kappa(A)$ is countable. Then*

$$\omega^{\kappa(A)} \leq o(A),$$

with equality if A is an integral domain. Moreover

$$O(A) = o(A) \cdot \omega = \omega^{\kappa(A)+1}.$$

We begin by showing that if A is an integral domain then there exists in A a descending chain of non zero ideals of type ω^κ, where $\kappa = \kappa(A)$. This is clear if A is a field. We assume therefore that $\kappa > 0$ and that, by Noetherian induction, the assertion is true for all proper quotients of A.

Case 1: There is a prime $\mathbf{p} \neq 0$ in A such that $\kappa = \kappa(A/\mathbf{p}) + 1$. For each $n > 0$, $\mathbf{p}^n/\mathbf{p}^{n+1}$ contains a submodule isomorphic to $A/\mathbf{p}$, and hence a descending chain of submodules of type $\omega^{\kappa(A/\mathbf{p})}$. Using these chains to refine the chain $\mathbf{p} \supset \mathbf{p}^2 \supset \mathbf{p}^3 \supset \ldots$ we obtain a chain of type $\omega^{\kappa(A/\mathbf{p})} \cdot \omega = \omega^{\kappa(A/\mathbf{p})+1} = \omega^\kappa$.

Case 2: There is a sequence $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \ldots$ of non zero prime ideals of A such that $\kappa = \sup_n \kappa(A/\mathbf{p}_n)$. (Since κ is assumed to be countable cases 1 and 2 exhaust all pos-

sibilities.) Put $J_n = \mathbf{p}_1 \dots \mathbf{p}_n$ $(n \geq 0)$. Then J_n/J_{n+1} contains a submodule isomorphic to $A/\mathbf{p}_{n+1}$, and hence a descending chain of submodules of type $\omega^{\kappa(A/\mathbf{p}_{n+1})}$. Using these chains to refine the chain $J_1 \supset J_2 \supset J_3 \dots$ we obtain a chain of type $\omega^{\kappa(A/\mathbf{p}_1)} + \omega^{\kappa(A/\mathbf{p}_2)} + \dots = \omega^{\sup \kappa(A/\mathbf{p}_n)} = \omega^{\kappa}$ (cf. [7, p. 309]).

Now that we have a chain of type ω^κ in A we can find one of type $\omega^\kappa + \dots + \omega^\kappa = (\omega^\kappa)\cdot n$ in A^n (cf. 1.4). Combining this observation with 2.11 we obtain the inequalities

$$(*) \qquad \omega^{\kappa+1} \leq \mathrm{O}(A) \leq \mathrm{o}(A)\cdot\omega .$$

We have been assuming A is an integral domain. We now drop this assumption and prove the theorem by Noetherian induction. The case $\kappa = 0$ (i.e. A is Artinian) is easily verified. Suppose therefore that $\kappa > 0$ and that the theorem holds for all proper quotients of A. Any finitely generated A-module M has a composition series with successive quotients of the form $A/\mathbf{p}$ for some prime ideal $\mathbf{p}$. It follows therefore as in the proof of 2.11 that $\mathrm{o}(M)$ is bounded above by a finite sum of terms of the form $\mathrm{o}(A/\mathbf{p})$. If we show that $\mathrm{o}(A/\mathbf{p}) \leq \omega^{\kappa(A/\mathbf{p})}$ then the reverse inequality follows from the conclusion established above for integral domains. We further obtain $\omega^{\kappa(A)} \leq \mathrm{o}(A) \leq \omega^{\kappa(A)}\cdot m$ for some integer $m \geq 1$ and so $\mathrm{o}(A)\cdot\omega = \omega^{\kappa(A)+1}$. In view of (*) above this implies that $\mathrm{O}(A) = \omega^{\kappa(A)+1}$ as well. Thus the proof of the theorem is reduced to the verification of: If A is an integral domain then $\mathrm{o}(A) \leq \omega^{\kappa(A)}$. Further we are permitted to assume the theorem is established for all proper quotients of A.

Let C be a descending chain of non zero ideals of A. For each $I \in C$ the chain C_I of ideals of C containing I corresponds to a chain in A/I and so $\mathrm{ord}(C_I) < \mathrm{O}(A/I) = \omega^{\kappa(A/I)+1} \leq \omega^{\kappa(A)}$. Hence $\mathrm{ord}(C) = \sup\{\mathrm{ord}(C_I) | I \in C\} \leq \omega^{\kappa(A)}$, so the theorem is proved.

2.13. **Remark**. Suppose $\kappa(A)$ is uncountable. Then since A has quotients of all possible (countable) Krull ordinals $< \kappa(A)$ it follows from Theorem 2.12 that $\mathrm{o}(A) \geq \Omega$. On the other hand Theorem 1.1 implies that $\mathrm{O}(A) \leq \Omega$. We conclude therefore that:

$$\kappa(A) \geq \Omega \Rightarrow \mathrm{o}(A) = \mathrm{O}(A) = \Omega .$$

Of course conjecture 2.9 predicts that this can never occur.

§3. The ordinal type of $S \Delta T$

Let A be a commutative Noetherian ring. If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of finitely generated modules then we have the relation $S(M) \leq S(M'') \Delta S(M')$ between the corresponding partially ordered sets of submodules. We shall prove here (3.7) that for any partially ordered sets S and T of ordinal types

$\leq \alpha$ we have $\mathrm{ord}\,(S \Delta T) \leq \alpha \cdot 3$. It follows therefore that $o(M) \leq \sup(o(M''),$ $o(M')) \cdot 3$. This is the fact used to conclude the proof of Lemma 2.11 above.

Though unnecessary for the applications above we shall give a fairly complete description of $\mathrm{ord}(S \Delta T)$. This result (Theorem 3.5) was worked out jointly with H. Hecht (Columbia University).

3.1. Additive ordinals. The following conditions on an ordinal σ are equivalent:

(1) $\alpha, \beta < \sigma \Rightarrow \alpha + \beta < \sigma$.

(2) $\sigma = \alpha + \beta \Rightarrow \sigma = \beta$.

(3) $\alpha < \sigma \Rightarrow \alpha \cdot \omega \leq \sigma$.

(4) $\alpha < \sigma \Rightarrow \alpha + \sigma = \sigma$.

(1) $\Rightarrow$ (3): If $\alpha < \sigma$ then (1) implies that $\alpha \cdot n < \sigma$ for all integers $n > 0$, and so $\sup_n \alpha \cdot n = \alpha \cdot \omega$ is $\leq \sigma$.

(3) $\Rightarrow$ (1): Say $\beta \leq \alpha < \sigma$. Then $\alpha + \beta \leq \alpha + \alpha < \alpha \cdot \omega \leq \sigma$, by (3).

The equivalence of (1), (2) and (4) is proved in Sierpinski [7, p. 279]. Sierpinski calls such ordinal numbers σ which are > 0, "prime components". We shall use the term "additive ordinal", which is suggested by (1): The segment $[\alpha)$ is stable under addition.

The importance of additive ordinals derives from the following theorem (see Sierpinski [7] Ch. XIV, §6):

3.2. Theorem. *Let α be an ordinal > 0.*

1) *There is an integer $n > 0$ and additive ordinals $\sigma_1 \geq \ldots \geq \sigma_n$ such that*

$$\alpha = \sigma_1 + \ldots + \sigma_n$$

2) *Both n and the additive ordinals σ_i above are unique. For example*

$$\sigma_1 = \sigma(\alpha)$$

is the greatest additive ordinal $\leq \alpha$.

3) *If $\alpha = \beta + \gamma$ with $\gamma > 0$ then there is a unique i, $1 \leq i \leq n$, such that $\gamma = \sigma_i + \ldots + \sigma_n$.*

It will be convenient for us to introduce the decomposition

$$(*) \qquad \alpha = \sigma(\alpha) \cdot q(\alpha) + \rho(\alpha)$$

where $q(\alpha)(> 0)$ denotes the number of σ_i equal to $\sigma_1 = \sigma(\alpha)$, and where $\rho(\alpha) = \sigma_{q(\alpha)+1} + \ldots + \sigma_n$. It is easily seen that $(*)$ is characterized by the properties

$\sigma(\alpha)$ is an additive ordinal,
$q(\alpha)$ is an integer > 0,
$\rho(\alpha) < \sigma(\alpha)$.

3.2. Proposition. *For any ordinal $\alpha > 0$, the least additive ordinal $> \alpha$ is*

$$\alpha \cdot \omega = \sigma(\alpha) \cdot \omega .$$

If σ is the least additive ordinal $> \alpha$ it follows from the relation $\sigma(\alpha) < \alpha < \sigma$ that $\sigma(\alpha) \cdot \omega \leq \sigma$. It remains to show that $\alpha \cdot \omega \leq \sigma(\alpha) \cdot \omega$ and that $\sigma(\alpha) \cdot \omega$ is additive. Since $\rho(\alpha) < \sigma(\alpha)$ we have $\rho(\alpha) + \sigma(\alpha) = \sigma(\alpha)$ and hence $\alpha \cdot n = \sigma(\alpha) \cdot (q(\alpha) \cdot n) + \rho(\alpha) < \sigma < \sigma(\alpha) \cdot (q(\alpha) \cdot n + 1) < \sigma(\alpha) \cdot \omega$. Taking the sup over n we obtain $\alpha \cdot \omega \leq \sigma(\alpha) \cdot \omega$.

If $\beta < \sigma(\alpha) \cdot \omega$ then $\beta < \sigma(\alpha) \cdot n$ for some n and so $\beta \cdot \omega \leq \sigma(\alpha) \cdot n \cdot \omega = \sigma(\alpha) \cdot \omega$. This shows that $\sigma(\alpha) \cdot \omega$ is additive, and thus concludes the proof.

3.4. The ordinal $\alpha \triangle \beta$. Let S and T be partially ordered sets. To compute $\operatorname{ord}(S \triangle T)$ we must consider well ordered subsets U of $S \triangle T$. The projection U_S (resp. U_T) of U in S (resp. T) is then also well ordered, and we have $U \subset U_S \triangle U_T$. It follows that

$$\operatorname{ord}(S \triangle T) = \operatorname{ord}(S' \triangle T')$$

where S' and T' vary over all well ordered subsets of S and T, respectively.

If α and β are ordinal numbers we shall put

$$\alpha \triangle \beta = \operatorname{ord}([\alpha] \triangle [\beta]).$$

Note that $\alpha \triangle \beta = \beta \triangle \alpha$

3.5. Theorem. *Let α and β be ordinal numbers such that $\alpha \geq \beta > 0$. Write $\alpha = \sigma(\alpha) \cdot q(\alpha) + \rho(\alpha)$ and $\beta = \sigma(\beta) \cdot q(\beta) + \rho(\beta)$ as in 3.2 (*) above. Then $\sigma(\alpha) \geq \sigma(\beta)$.*

If $\sigma(\alpha) > \sigma(\beta)$ then $\alpha \triangle \beta = \sigma(\alpha) \cdot q(\alpha) + \rho(\alpha) \triangle \beta$. If $\sigma(\alpha) = \sigma(\beta)$ then

$$\alpha \triangle \beta = \begin{cases} \sigma(\alpha)\cdot(q(\alpha) + q(\beta) - 1) & \text{if } \rho(\alpha) = \rho(\beta) = 0 \\ \sigma(\alpha)\cdot(q(\alpha) + q(\beta)) + \rho(\alpha) \triangle \rho(\beta) & \text{otherwise.} \end{cases}$$

Before giving the proof we shall draw some consequences.

3.6. Corollary. *With the notation of Theorem 3.5 we have:*

$$\sigma(\alpha \triangle \beta) = \sigma(\alpha).$$

$$q(\alpha \triangle \beta) \leq 2q(\alpha) .$$

$$\alpha \leq \alpha \triangle \beta \leq \alpha + \beta + \alpha \leq \alpha \cdot 3 .$$

$$(\alpha \triangle \beta) \cdot \omega = \alpha \cdot \omega .$$

The first two assertions follow by induction on α, using the fact that when

$\sigma(\alpha) = \sigma(\beta)$ we have $q(\alpha) \geq q(\beta)$. (Otherwise the condition $\alpha > \beta$ would be violated.)

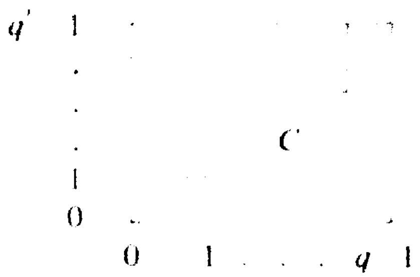
The last assertions follow immediately from the first with the aid of 3.3.

3.7. Corollary. *Let S and T be non empty partially ordered sets such that* $\operatorname{ord}(S) \geq \operatorname{ord}(T)$. *Then* $\operatorname{ord} S \leq \operatorname{ord}(S \Delta T) < \operatorname{ord} S + \operatorname{ord} T + \operatorname{ord} S \leq (\operatorname{ord} S) \cdot 3$.

If $\operatorname{ord} T = 0$, i.e. if no two elements of T are comparable, it is clear that $\operatorname{ord}(S \Delta T) = \operatorname{ord} S$. If $\operatorname{ord} T > 0$ then $\operatorname{ord}(S \Delta T) = \sup \operatorname{ord}(S' \Delta T')$ where S' and T' vary over well ordered subsets of S and T, respectively (see 3.4), and there is no loss in restricting to pairs (S', T') for which $\operatorname{ord} S' \geq \operatorname{ord} T' > 0$. We then obtain the corollary by applying Corollary 3.6 to estimate each $\operatorname{ord}(S' \Delta T')$ and then taking the supremum.

3.8. Proof of Theorem 3.5. We argue by induction on $\sigma = \sigma(\alpha)$.

$\sigma = 1$. Then $\alpha = q$ and $\beta = q'$ are finite and one sees easily that any maximal chain in $[q) \times [q') = \{0, 1, \ldots, q - 1\} \times \{0, 1, \ldots, q' - 1\}$ has length $q + q' - 1$.



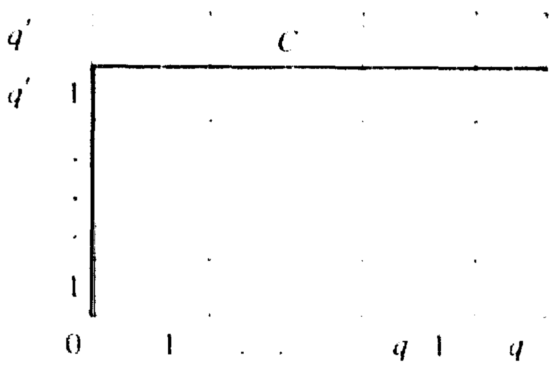
$\sigma > 1$. We begin by showing that:

a) $\sigma \Delta \sigma = \sigma$.

For let $S = [\sigma)$ and let C be a well ordered subset of $S \Delta S$. If $c \in C$ the initial segment C_c preceding c in C lies in $S_1 \Delta S_2$ for suitable initial segments S_1 and S_2 of S. By induction we can apply the theorem and Corollary 3.6 to $\operatorname{ord} S_1$ and $\operatorname{ord} S_2$, which are $< \sigma$. It follows that $\operatorname{ord} C_c$ is $< \sup(\operatorname{ord} S_1, \operatorname{ord} S_2) \cdot 3 < \sigma$. Thus $\operatorname{ord} C = \sup_{c \in C} \operatorname{ord} C_c \leq \sigma$. Taking for C the diagonal in $S \Delta S$ we have an example with $\operatorname{ord} C = \sigma$. Thus $\sigma \Delta \sigma = \sigma$, as claimed.

We next claim:

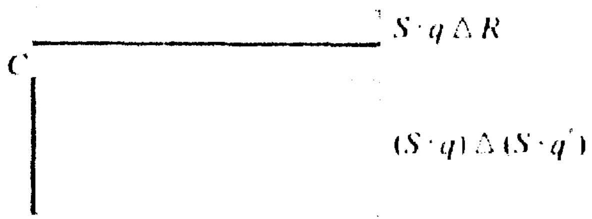
b) If q and q' are integers > 0 then $(\sigma \cdot q) \Delta (\sigma \cdot q') = \sigma \cdot (q + q' - 1)$.



With $S = [\sigma)$ as above let $S \cdot q$ denote $S \overset{<}{+} \ldots \overset{<}{+} S$ (q terms) (see 1.2) and similarly for $S \cdot q'$. Then a chain C in $(S \cdot q) \triangle (S \cdot q')$ can meet at most $q + q' - 1$ of the qq' "squares" of the form $S \triangle S$ (cf. the proof above in the case $\sigma = 1$). Hence $\operatorname{ord} C \leq (\operatorname{ord}(S \triangle S)) \cdot (q + q' - 1) = \sigma \cdot (q + q' - 1)$, by a). On the other hand the totally ordered set $C = (\{0\} \times S \cdot (q' - 1)) \cup (S \cdot q \times \{0'\})$, where 0 is the initial element of the first copy of S in $S \cdot q$, and where $0'$ is the initial element of the q'th copy of S in $S \cdot q'$, is of type $\sigma \cdot (q + q' - 1)$; whence b).

Next suppose that $0 < \rho' < \sigma$. We claim that:

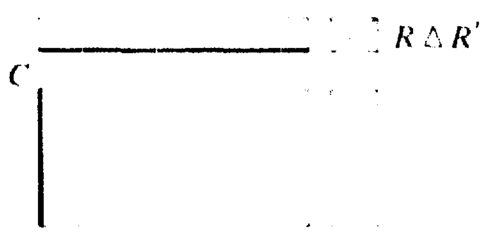
c) $(\sigma \cdot q) \triangle (\sigma \cdot q' + \rho') = \sigma \cdot (q + q')$.



If r' is the initial element of $R' = [\rho')$ then $C = (\{0\} \times S \cdot q') \cup (\{r'\} \times S \cdot q)$ is a well ordered set of type $\sigma \cdot (q + q')$ in $(S \cdot q) \triangle (S \cdot q' \overset{<}{+} R)$, so the left side in c) dominates the right side. On the other hand the left side is $\leq (\sigma \cdot q) \triangle (\sigma \cdot (q' + 1))$ which, by b), is $\leq \sigma \cdot (q + q')$; whence c).

Assume now that q, q' are integers > 0 and that ρ, ρ' are ordinals > 0 and $< \sigma$. We claim

d) $(\sigma \cdot q + \rho) \triangle (\sigma \cdot q' + \rho') = \sigma \cdot (q + q') + \rho \triangle \rho'$



Keep the above notation and let r denote the initial element of $R = [\rho)$. In $(S \cdot q \leq + R) \Delta (S \cdot q' \leq + R')$ the set $C = (\{0\} \times S \cdot q') \cup (S \cdot q \times \{r'\})$ is well ordered of type $\sigma \cdot (q + q')$, and it entirely precedes $R \Delta R'$. It follows that the left side of d) dominates the right side. For the reverse inequality consider any chain D in $(S \cdot q < + R) \Delta (S \cdot q' < + R')$. Write $D = D_0 < + D_1$ where $D_1 = D \cap (R \Delta R')$. Then D_0 is contained in either $(S \cdot q < + R) \Delta (S \cdot q')$ or in $(S \cdot q) \Delta (S \cdot q' \leq + R')$. In either case it follows from c) that $\operatorname{ord} D_0 \leq \sigma \cdot (q + q')$. Hence $\operatorname{ord} D = \operatorname{ord} D_0 + \operatorname{ord} D_1 \leq \sigma \cdot (q + q') + \rho \Delta \rho'$, whence $d)$.

To conclude the proof we show that, for any ordinal β such that $0 \leq \beta \leq \sigma$ we have

e) $(\sigma \cdot q + \rho) \Delta \beta = \sigma \cdot p + \rho \Delta \beta$

If b is the initial element of $B = [\beta)$ then $S \cdot q \times \{b\}$ is a chain of type $\sigma \cdot q$ in $(S \cdot q \leq + R) \Delta B$ entirely preceding $R \Delta B$, and so $\sigma \cdot q + \rho \Delta \beta \leq (\sigma \cdot q + \rho) \Delta \beta$. Let D be any chain in $(S \cdot q \leq + R) \Delta B$ and write $D = D_0 \leq + D_1$ with $D_1 = D \cap (R \Delta B)$ and $D_0 = D \cap (S \cdot q \Delta B)$. Since $\operatorname{ord} D_1 \leq \rho \Delta \beta$ it will suffice, in order to show $\operatorname{ord} D \leq \sigma \cdot q + \rho \Delta \beta$, to show that $\operatorname{ord} D_0 \leq \sigma \cdot q$, or to show that $(\sigma \cdot q) \Delta \beta \leq \sigma \cdot q$. But $(\sigma \cdot q) \Delta \sigma = \sigma \cdot q$ by b).

This concludes the proof of Theorem 3.5.

Note added in proof: G. Sabbagh has pointed out to the author that the essential content of Theorem 3.5 has been treated in Milner and Rado, Proc. Lond. Math. Soc. 15 (1965) 750–768, especially Lemma 2 on p. 760, which is due to Toulmin.

References

[1] N. Bourbaki, Algèbre commutative, Ch. 4 Indeaux premiers associés et décomposition primaire (Hermann, Paris, 1961).

[2] P. Eklof and G. Sabbagh, Definability problems for modules and rings, preprint, Yale Univ. (1970).

[3] G. Krause, On the Krull-dimension of left Noetherian left Matlis-rings, Math. Z. 118 (1970) 207–214.

[4] W. Krull, Dimensionstheorie in Stellenringen, J. Reine Angew. Math. 179 (1938) 204–226.

[5] M. Negata, Local rings (Interscience, 1962).

[6] J.E. Schneider, Descending chains of ideals in Noetherian rings I, preprint, Penn. State Univ. (1971).

[7] W. Sierpinski, Cardinal and ordinal numbers, Polska Akad. Naut., Monog. Matem., Warsaw (1958).

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 4 (1971) pp. 361–372

A COHOMOLOGICAL CRITERION FOR p-NILPOTENCE

Daniel QUILLEN *

Dept. of Mathematics, Mass. Institute of Technology, Cambridge, Mass. 02139, U.S.A.

Received 12 May 1971

I present here a theorem asserting that a finite group has a normal p-complement provided p is odd and the varieties of the cohomology rings of the group and its Sylow p-subgroup are isomorphic. Although the result perhaps is not too useful, its proof serves to illustrate the theorems of [3] on the structure of cohomology rings. In addition I describe an alternative approach to these results in the case of finite groups using the norm mapping of Evens [6].

§ 1. Statement of the theorem

Let p be a prime number, let G be a finite group and let P be a Sylow p-subgroup of G. Denote by $H^*(G)$ the cohomology ring of G with coefficients in the ring $\mathbf{Z}/p\mathbf{Z}$ with trivial G-action. One knows by the theory of the transfer that the restriction homomorphism

(1.1) $\qquad H^i(G) \longrightarrow H^i(P)$

is injective for all i.

Recall that one says that G is p-nilpotent, or that G has a normal p-complement, if the p'-elements (order prime to p) of G form a subgroup. If N is this subgroup, then clearly N is normal and the composition $P \to G \to G/N$ is an isomorphism. It follows that the restriction homomorphism 1.1 is an isomorphism in this case.

We shall be interested in results going in the converse direction, that is which establish p-nilpotence from some hypothesis on the restriction homomorphism. The simplest of these is the following.

* Supported by the National Science Foundation.

Theorem 1.2. (Tate [4].) *If $H^1(G) \simeq H^1(P)$, then G is p-nilpotent.*

In contrast we have the following two results making use of the cohomology in large degrees.

Theorem 1.3. (Atiyah) *If $H^i(G) \simeq H^i(P)$ for all sufficiently large i, then G is p-nilpotent.*

When p is odd, Atiyah's theorem evidently is a consequence of the following, whose proof is the goal of this paper.

Theorem 1.4. *Assume that for each cohomology class $x \in H^*(P)$ of even degree there is a power q of p such that x^q is the restriction of a class in $H^*(G)$. If p is odd, then G is p-nilpotent.*

Here is an example showing why p must be odd. Let P be the quaternion group of order 8 and let G be the semi-direct product of P and a cyclic group of order 3 where the generator of the cyclic group cyclically permutes i, j, k. The mod 2 cohomology of P is known to be

$$H^*(P) = (\mathbf{Z}/2\mathbf{Z})[x, y, e]/(x^2 + xy + y^2, x^2y + xy^2)$$

where x, y is a basis for $H^1(P)$ and e is the unique non-zero element of $H^4(P)$. For $i \equiv 1, 2 \pmod 4$, $H^i(P)$ is the two-dimensional irreducible representation of $\mathbf{Z}/3\mathbf{Z}$, and in the other dimensions it acts trivially, so

$$H^*(G) = H^*(P)^{\mathbf{Z}/3\mathbf{Z}} = (\mathbf{Z}/2\mathbf{Z})[z, e]$$

where z and e are the non-zero elements of degrees 3 and 4 respectively. As x, y are nilpotent one sees that the first hypothesis of 1.4 holds, but G is not p-nilpotent.

Since 1.4 does not cover the case $p = 2$ of 1.3 we sketch the proof of Atiyah's theorem for the convenience of the reader. One uses the spectral sequence [1] which starts with the integral cohomology ring $H^*(G, \mathbf{Z})$ and abuts to the completion $R(G)^\wedge$ of the character ring with respect to the $\bar{R}(G)$-adic topology, where $\bar{R}(G)$ is the augmentation ideal. One needs also the fact that the p-primary part $\bar{R}(G)^\wedge_p$ of the completion of $R(G)$ is a free module over the p-adic integers of rank equal to the number of conjugacy classes of non-identity p-elements (order a power of p) in G. Now the hypothesis of 1.3 implies that the restriction from $H^i(G, \mathbf{Z})$ to $H^i(P, \mathbf{Z})$ induces an isomorphism on the p-primary components for all sufficiently large i. By means of the spectral sequence one can deduce from this that the kernel and cokernel of the restriction homomorphism from $R(G)^\wedge_p$ to $R(P)^\wedge_p$ are finite. It follows that G and P have the same number of conjugacy classes of p-elements, that is, there is no fusion of elements of P in G. Now it is difficult to show this implies G is p-nilpotent [2, IV, 4.9].

The proof of 1.4 is similar in spirit, namely by use of various results in the structure of the ring $H^*(G)$ it reduces to the purely group-theoretical result below, which, however, is not as elementary as the one used in Atiyah's theorem.

Denote by $N_G(A)$ (resp. $C_G(A)$) the subgroup of G consisting of elements normalizing (resp. centralizing) A, and recall that when $A \subseteq P \subseteq G$, A is said to be weakly-closed in P with respect to G if $g^{-1}Ag \subseteq P$ implies $g^{-1}Ag = A$.

Theorem 1.5. *Assume for every normal elementary abelian p-subgroup A of P that A is weakly closed in P with respect to G and that $N_G(A)/C_G(A)$ is a p-group. If p is odd, then G is p-nilpotent.*

The above example shows that it is necessary for p to be odd.

The reduction of 1.4 to 1.5 appears in the next section. We show in the third section how the techniques of Section 2 can be used to simplify somewhat the results of [3] in the case of finite groups. Finally the last section contains the proof of 1.5.

§2. The norm

Throughout this paper we work with finite groups and use letters A, A', etc. to denote elementary abelian p-groups. If $K \subseteq G$ and $u \in H^*(G)$, we let $u \mid K$ denote the image of u under the restriction homomorphism from $H^*(G)$ to $H^*(K)$. Define a graded commutative ring $H(G)$ by

$$H(G) = \begin{cases} \bigoplus_{i \geq 0} H^{2i}(G) & p \text{ odd} \\ \bigoplus_{i \geq 0} H^{i}(G) & p = 2\,; \end{cases}$$

it is finitely generated algebra over $\mathbf{Z}/p\mathbf{Z}$ (cf. [3, §2]).

If G' is a subgroup of G, let

$$\operatorname{Norm}_{G' \to G} : H(G') \to H(G)$$

be the multiplicative analogue of the transfer introduced by Evens [6]. This operation satisfies the double coset formula [6, Prop. 2]

$$(\operatorname{Norm}_{G' \to G} u) \mid K = \prod_{G'gK} \operatorname{Norm}_{K \cap g^{-1}G'g \to K}(i_g^* u)$$

where g runs over a system of representatives for the (G', K)-double cosets and where i_g is the homomorphism from $K \cap g^{-1}G'g$ to G' given by $i_g(x) = gxg^{-1}$.

Consequently if $u \in H(G')$ is such that $u|G'' = 1$ for all $G'' < G'$, then we have

$$(2.1)\qquad (\mathrm{Norm}_{G' \to G} u)|K = \begin{cases} 1 & \text{if } G' \nrightarrow K \\ \prod_{g \in I} i_g^* u & \text{if } K = G' \end{cases}$$

where I is a system of coset representatives for G' in its normalizer $N_G(G')$, and where the notation $G' \nrightarrow K$ means that G' is not conjugate to a subgroup of K.

Given an elementary abelian p-group A, we identify $H^1(A)$ with the dual $A^{\#}$ of A as a vector space over $\mathbf{Z}/p\mathbf{Z}$ by means of the canonical isomorphism. We recall that the Bockstein homomorphism

$$(2.2)\qquad \beta: A^{\#} \to H^2(A)$$

is injective and that there are canonical isomorphisms

$$(2.3)\qquad H^*(A) = \begin{cases} \Lambda(A^{\#}) \otimes S(\beta A^{\#}) & p \text{ odd} \\ S(A^{\#}) & p = 2\,, \end{cases}$$

where Λ and S are respectively the exterior and symmetric algebra functors on vector spaces over $\mathbf{Z}/p\mathbf{Z}$. Indeed these facts are well-known when A is cyclic and follow in general by the Kunneth formula for the cohomology of a product of two groups.

Set

$$e_A = \prod_{0 \neq u \in A^{\#}} \beta u \in H^d(A)$$

where $d = 2p^r - 2$ and r is the rank of A. It is clear that e_A is a non-zero-divisor in $H(A)$ such that $e_A|A' = 0$ for all $A' < A$, and such that $\theta^* e_A = e_A$ for all automorphisms θ of A.

Now let A be an elementary abelian p-subgroup of G.

Lemma 2.4. *Let $[N_G(A):A] = qh$ where $(h, p) = 1$ and q is a power of p. Then there exists v_A in $H(G)$ such that*

$$v_A|A' = \begin{cases} 0 & \text{if } A \nrightarrow A' \\ e_A^q & \text{if } A = A'. \end{cases}$$

Moreover if $v \in H(A)$ is invariant under $N_G(A)$, then there is a $\alpha(v)$ in $H(G)$ with $\alpha(v)|A = v^q e_A^q$.

Proof. Set $z = \mathrm{Norm}_{A \to G}(1 + e_A)$. Formula 2.1 shows that $z|A' = 1$ if $A \not\to A'$ and that

$$z|A = (1 + e_A)^{qh} = (1 + e_A^q)^h$$
$$= 1 + he_A^q + \text{terms of higher degree}.$$

Thus if v_A is taken to be $(1/h)$-times the homogeneous component of z of degree $q \cdot \deg e_A$, it has the required properties. To prove the existence of $\alpha(y)$ we can obviously suppose that y is homogeneous. If $z' = \mathrm{Norm}_{A \to G}(1 + ye_A)$, then 2.1 shows that

$$z'|A = 1 + hy^q e_A^q + \text{terms of higher degree}$$

so we can take $\alpha(y)$ to be $(1/h)$-times the homogeneous component of z' of degree $q(\deg y + \deg e_A)$. q.e.d.

From 2.3 it is clear that

(2.5) $\qquad H(A)/\sqrt{0} \sim S(A^{\#})$

where $\sqrt{0}$ denotes the ideal of nilpotent elements. When $A \subset G$ we let $\mathfrak{p}_A$ denote the ideal in $H(G)$ consisting of the elements u such that $u|A$ is nilpotent. Then $\mathfrak{p}_A$ is a prime ideal because the restriction to A induces an injection

(2.6) $\qquad H(G)/\mathfrak{p}_A \hookrightarrow S(A^{\#})$.

Theorem 2.7. *Given $A, A' \subset G$, then $\mathfrak{p}_A \supset \mathfrak{p}_{A'}$ if and only if A is conjugate to a subgroup of A'.*

Proof. The "if" part is clear when A is a subgroup of A' and the general case reduces to this one because inner automorphisms of a group act trivially on its cohomology. Conversely suppose that $A \not\to A'$; by the lemma $v_A \in \mathfrak{p}_{A'}$, yet $v_A \notin \mathfrak{p}_A$ as the image of e_A in $S(A^{\#})$ is non-zero, so the theorem is proved.

Denote by

(2.8) $\qquad k(\mathfrak{p}_A) \hookrightarrow k(A)$

the extension of quotient fields associated to the homomorphism 2.6. The group

(2.9) $\qquad W = N_G(A)/C_G(A)$

acts as automorphisms of this field extension in a natural way.

Theorem 2.10. *The extension 2.8 is a finite normal extension and W is isomorphic to its group of automorphisms.*

Proof. The extension is finite because quite generally the cohomology ring of a subgroup is a finitely generated module over the cohomology ring of the group (cf. [3, §2]). Given an element z of $S(A^{\#})$ invariant under the action of W, there is an invariant element v of $H(A)$ corresponding to z under the isomorphism 2.5, and by Lemma 2.4 there is an element $\alpha(v)$ of $H(G)$ restricting to $v^q e_A^q$ where q is a power of p. Consequently the homomorphism

$$(H(G)/\mathfrak{p}_A)[v_A^{-1}] \hookrightarrow (S(A^{\#})[e_A^{-1}])^W$$

obtained by localizing 2.6 with respect to the multiplicative system of powers of v_A has the property that the qth power of any element of its target is in its image. Therefore the extension 2.8 is a composite

$$k(\mathfrak{p}_A) \hookrightarrow k(A)^W \hookrightarrow k(A)$$

where the first extension is purely inseparable and the second is Galoisian with group W, as W acts faithfully on $k(A)$. The theorem follows.

We now show why Theorem 1.4 follows from 1.5. Identify $H(G)$ with a subring of $H(P)$ by means of the restriction homomorphism, and denote by $\mathfrak{p}_A \subset H(P)$ (resp. $\mathfrak{q}_A \subset H(G)$) the prime ideal associated to an elementary abelian p-subgroup A of P (resp. A' of G). Clearly $\mathfrak{q}_A = \mathfrak{p}_A \cap H(G)$. Assuming the hypotheses of 1.4 hold, any element of $H(P)$ lies in $H(G)$ after being raised to some power of p, hence there is a one-one correspondence between prime ideals $\mathfrak{p}$ in $H(P)$ and $\mathfrak{q}$ of $H(G)$ given by the formulas

$$\mathfrak{q} = \mathfrak{p} \cap H(G)$$

$$\mathfrak{p} = \{x \in H(P) \mid x^{p^n} \in \mathfrak{q} \text{ for some } n\}.$$

If $A, A' \subset P$ become conjugate in G, then $\mathfrak{q}_A = \mathfrak{q}_{A'}$, so $\mathfrak{p}_A = \mathfrak{p}_{A'}$, and by Theorem 2.7 the subgroups A and A' are conjugate in P. Thus there is no fusion of elementary abelian p-subgroups of P in G; in particular, if A is normal in P, then A is weakly closed in P with respect to G. Finally for any A in P we have that $k(\mathfrak{p}_A)$ is a purely inseparable extension of $k(\mathfrak{q}_A)$, hence using 2.10

$$N_P(A)/C_P(A) = \mathrm{Aut}(k(A)/k(\mathfrak{p}_A)) = \mathrm{Aut}(k(A)/k(\mathfrak{q}_A)) = N_G(A)/C_G(A).$$

In particular $N_G(A)/C_G(A)$ is a p-group, showing that the hypotheses of 1.5 hold.

§3. Properties of the ring $H(G)$

The material of the following section is not required for the proof of Theorem 1.4. Its purpose is to show how the arguments of §2 using the norm mapping of Evens furnish an alternative approach to the results of [3] in the case of cohomology rings of finite groups. It is hoped that someone mainly interested in this case will find the version here easier to understand than the one in [3] where the theorems are formulated for the equivariant cohomology rings of G-spaces. However even for the cohomology rings of finite groups some consideration of equivariant cohomology is necessary, e.g. the proof of 3.1 below.

The following theorem is half of the main theorem of [3]; its proof is deferred to the end of this section.

Theorem 3.1. *If $u \in H^q(G)$ is such that $u|A = 0$ for all elementary abelian p-subgroups A of G, then u is nilpotent.*

Corollary 3.2. *The minimal prime ideals of $H(G)$ correspond bijectively to the conjugacy classes of maximal elementary abelian p-subgroups of G by the rule $A \longleftrightarrow \mathfrak{p}_A$.*

Proof. If A_i, $1 \leqslant i \leqslant r$ are representatives for the conjugacy classes of maximal $A \subset G$, then by the theorem we have that the ideal of nilpotent elements of $H(G)$ admits a representation

$$\sqrt{0} = \bigcap_{i=1}^{r} \mathfrak{p}_{A_i}.$$

On the other hand this representation is irredundant by 2.7, so the $\mathfrak{p}_{A_i}$ are the minimal prime ideals of $H(G)$.

Corollary 3.3. *The Krull dimension of $H(G)$ equals the maximum rank of an elementary abelian p-subgroup of G.*

Proof. One of the many characterizations of the Krull dimension of a finitely generated commutative algebra over a field is as the maximum of the transcendence degrees of the residue fields associated to the different minimal prime ideals. By 2.10 the transcendence degree of $k(\mathfrak{p}_A)$ over $\mathbf{Z}/p\mathbf{Z}$ coincides with that of $k(A)$, and $k(A)$ has transcendence degree $r = \operatorname{rank} A$ because it is the quotient field of $S(A^{\#})$ which is isomorphic to a polynomial ring with r generators. Since all the minimal prime ideals of $H(G)$ are of the form $\mathfrak{p}_A$ with A maximal by 3.2, the corollary follows.

We choose an algebraically closed field Ω of characteristic p and form the *variety* $H(G)(\Omega)$ of the ring $H(G)$; it is the set of ring homomorphisms from $H(G)$ endowed

with the Zariski topology. For an elementary abelian p-group one has

$$H(A)(\Omega) \simeq S(A^{\#})(\Omega) \simeq A \otimes \Omega, \tag{3.4}$$

where the tensor product is taken over $\mathbf{Z}/p\mathbf{Z}$, hence the variety of $H(A)$ is an affine space over Ω of dimension equal to the rank of A. If $A \subset G$ the common zeroes of the elements of $\mathfrak{p}_A$ form an irreducible closed subspace of $H(G)(\Omega)$ which will be denoted V_A. This subspace may also be described as the image of the map

$$A_* : A \otimes \Omega \to H(G)(\Omega) \tag{3.5}$$

induced by the restriction homomorphism from $H(G)$ to $H(A)$, because on one hand the injectivity of 2.6 implies that V_A is the closure of the image and on the other hand the map is closed by the Cohen–Seidenberg theorem, as $H(A)$ is a finitely generated $H(G)$-module.

Set

$$V_A^+ = V_A - \bigcup_{A' < A} V_{A'}$$

$$(A \otimes \Omega)^+ = A \otimes \Omega - \bigcup_{A' < A} A' \otimes \Omega .$$

Lemma 3.6. *$(A \otimes \Omega)^+$ is the Zariski open subset of $A \otimes \Omega$ on which e_A is invertible. V_A^+ is the open subset of V_A on which the element v_A is invertible (notation as in 2.4). Furthermore $A_*^{-1} V_A^+ = (A \otimes \Omega)^+$.*

Proof. The first statement is clear and so is the inclusion $A_*^{-1} V_A^+ \subset (A \otimes \Omega)^+$. If U is the subset of V_A on which v_A is invertible, then $U \subset V_A^+$ since $v_A \in \mathfrak{p}_{A'}$ for $A' < A$. But v_A restricts to e_A^q, hence $A_*^{-1} U = (A \otimes \Omega)^+$. Thus we have $A_*^{-1} V_A^+ = (A \otimes \Omega)^+$, hence also $U = V_A^+$ because A_* maps $A \otimes \Omega$ onto V_A. q.e.d.

It follows from the lemma that V_A^+ is an affine variety with coordinate ring $(H(G)/\mathfrak{p}_A)[v_A^{-1}]$. But we have seen in the proof of 2.10 that the homomorphism induced by the restriction to A

$$(H(G)/\mathfrak{p}_A)[v_A^{-1}] \to (S(A^{\#})[e_A^{-1}])^W$$

is injective and that any elements of its target when raised to a power of p lies in its image. Consequently this homomorphism induces a homeomorphism of varieties. Because the variety of the subring of invariants for the action of a finite group on a ring is the quotient space of the variety of the ring by the group, A_* induces a homeomorphism of V_A^+ with the quotient of $(A \otimes \Omega)^+$ by W.

Next we note that V_A^+ and $V_{A'}^+$ are disjoint if A and A' are not conjugate, because if say A is not conjugate to a subgroup of A', then v_A is invertible on V_A^+ and zero on $V_{A'}^+$. Finally $H(G)(\Omega)$ is the union of the V_A^+ as the irreducible components of the variety are of the form V_A with A maximal by 3.2. Therefore we have proved the following *stratification theorem:*

Theorem 3.7. *If I is a set of representatives for the conjugacy classes of elementary abelian p-subgroups of G, then the variety of $H(G)$ admits a stratification*

$$H(G)(\Omega) = \bigcup_{A \in I} V_A^+$$

into disjoint locally closed irreducible affine subvarieties. Moreover the group $W_G(A) = N_G(A)/C_G(A)$ acts freely on $(A \otimes \Omega)^+$ and the restriction homomorphism from $H(G)$ to $H(A)$ induces a homeomorphism

$$(A \otimes \Omega)^+ / W_G(A) \sim V_A^+ .$$

We finish this section with the proof of 3.1. First consider the case where G is abelian and let $A \subset G$ be the subgroup of elements of order 1 or p. If L is the kernel of the restriction from $H^1(G)$ to $H^1(A)$, then the kernel J of the restriction homomorphism from $H^*(G)$ to $H^*(A)$ is the ideal generated by L. Indeed if G is cyclic this is easy to see, and in the general case it follows by writing G as a product of cyclic groups and using the Kunneth formula. The square of any element of L is zero by anti-commutativity for p odd and for $p = 2$ it can be proved by reduction to the cyclic case. Consequently the ideal J consists of nilpotent elements, in fact $z^p = 0$ for all z in J, so we see that 3.1 is true when G is abelian.

To reduce the theorem to this special case we use equivariant cohomology. Let BG be a classifying space for G, that is, a connected CW complex with fundamental group G whose universal covering PG is contractible. Then PG is a principal G-bundle over BG with G acting as deck translations, so if X is a space on which G acts we can form the orbit space $PG \times^G X$ of the action $g(p, x) = (gp, gx)$ on $PG \times X$. The equivariant cohomology of the G-space X is defined by the formula

$$H_G^*(X) = H^*(PG \times^G X)$$

and one can show that it is independent of the choice of BG and that it behaves functorially in the pair (G, X) (cf. [3, §1]). When $X = pt$, the space consisting of a point, the equivariant cohomology of X is simply the cohomology of the group.

$$H_G^*(pt) \simeq H^*(G).$$

Lemma 3.8. *If G' is a subgroup of G, then the homomorphism*

$$H^*_G(pt) \to H^*_G(G/G')$$

induced by the map of G/G' to a point is isomorphic to the restriction homomorphism from $H^*(G)$ to $H^*(G')$.

Proof. Since G' acts freely on PG, we can take for BG' the orbit space PG/G', whence the restriction homomorphism from $H^*(G)$ to $H^*(G')$ is isomorphic to the map on cohomology induced by the projection of PG/G' onto BG. As $PG \times^G (G/G') \sim PG/G'$, the lemma follows.

In the lemma below we assume for simplicity that X is a compact C^∞ manifold on which G acts by diffeomorphisms (for a more general version see [3, 3.2]).

Lemma 3.9. *If $u \in H^*_G(X)$ restricts to zero on each orbit of X, then u is nilpotent.*

Proof. We first note that any orbit Gx has a G-invariant neighborhood U such that $H^*_G(U) \sim H^*_G(Gx)$. In effect by averaging we can produce an equivariant Riemannian structure on X and take U to be a small tubular neighborhood of Gx with respect to this metric. Then U will be stable under G and the inclusion of Gx in U will be a G-homotopy equivalence, i.e. the homotopy inverse and homotopies will be G-equivariant, hence the inclusion will induce an isomorphism on equivariant cohomology.

Thus each orbit has an invariant neighborhood U such that the restriction $u \mid U$ of u to U is zero. By compactness X can be covered by invariant open sets $U_1, \ldots, U_n$ such that $u \mid U_i = 0$ for each i. If $V_m = U_1 \cup \ldots \cup U_m$ we show by induction on m that $u^m \mid V_m = 0$, this being clear for $m = 1$. Assuming true for $m-1$, we consider the Mayer–Vietoris sequence

$$H^{*-1}_G(U_m \cap V_{m-1}) \xrightarrow{\delta} H^*_G(V_m) \longrightarrow H^*_G(U_m) \oplus H^*_G(V_{m-1}).$$

Then the element u^{m-1} restricts to zero on U_m and on V_{m-1} by induction hypothesis, hence $u^{m-1} \mid V_m = \delta z$ for some z in $H^{*-1}_G(U_m \cap V_{m-1})$. As δ is a homomorphism of $H^*_G(X)$-modules,

$$u^m \mid V_m = \delta z \cdot u = \delta(zu) = 0$$

as u restricts to zero on the intersection. This completes the induction, so taking $m = n$ we have that $u^n = 0$ in $H^*_G(X)$, proving the lemma.

We now finish the proof of 3.1. Let $u \in H^*(G)$ be such that $u \mid A = 0$ for all $A \subset G$. If G' is an abelian subgroup of G, then $u \mid G'$ is nilpotent as the theorem is

is proved for abelian groups; replacing u by u^n for some n, we can assume that $u \mid G' = 0$ for all abelian subgroups G'.

Let V be a faithful finite-dimensional unitary representation of G and identify G with its image in the unitary group U of V. Let G act by left multiplication on the flag manifold $X = U/T$, where T is a maximal torus of U, and let v be the image of u under the map equivariant cohomology

$$(3.10) \qquad H^*(G) \sim H^*_G(pt) \longrightarrow H^*_G(X)$$

induced by the map from X to a point. The orbits of G on X are of the form G/G', where G' is abelian as it is conjugate in U to a subgroup of T. By 3.8, v restricts to zero on each orbit of X, hence v is nilpotent by 3.9. But the homomorphism 3.10 is *injective*. Indeed $PG \times^G X$ may be identified with the bundle of flags in the vector bundle $PG \times^G V$ over BG, and one knows by the so-called splitting principle in the theory of Chern classes that the induced map of cohomology from the base to such a flag bundle is injective. Therefore we conclude that u is nilpotent and the proof is complete.

§4. Proof of 1.5

We need the following general fact about elementary abelian p-subgroups

Proposition 4.1. *Let P be a Sylow p-subgroup of the finite group G and let A be a maximal normal elementary abelian subgroup of P. If p is odd, then A is already a maximal elementary abelian p-subgroup of G, that is, A is the set of elements of order 1 or p of $C_G(A)$.*

For a proof that A is a maximal elementary abelian subgroup of P, see [2, III, 12.1]. If $A' \subseteq G$ contains A, then $A' \subseteq N_G(A)$, and as P is a Sylow p-subgroup of $N_G(A)$, one can conjugate A' into a subgroup of P by an element normalizing A. By the maximality of A in P we have $A = A'$, proving the proposition.

We now give the proof of 1.5. Consider first the case in which every element of order p of G is in the center $Z(G)$. (The counterexample for $p = 2$ is of this form.) Then G is p-nilpotent by [2, IV, 5.5], or alternatively by the well-known criterion of Frobenius and the fact that for p odd a p'-automorphism of a p-group is trivial if it acts trivially on the elements of order p.

Now suppose 1.5 is false, let G be a counterexample of least order and choose a maximal normal elementary abelian subgroup A of P. We are going to show using the minimality of G that A is normal in G. Assuming this, the hypotheses of 1.5 imply that $G/C_G(A)$ is a p-group, hence $C_G(A)$ is not p-nilpotent or else G would be also. But by 4.1 the group $C_G(A)$ has the property that every element of order p

is in the center, hence this group is p-nilpotent by the case of the theorem already proved. This contradiction then establishes the theorem.

So let A be a normal elementary abelian subgroup of P and suppose that A is not normal in G. Let B be a maximal G-normal subgroup of A and define the subgroups A_1 and N by

$$A_1/B = A/B \cap Z(P/B)$$

$$N = N_G(A_1).$$

Since A/B is a non-trivial normal subgroup of the p-group P/B, it intersects the center non-trivially, so $A_1 > B$ and $N < G$ by the maximality of B. Since A_1 is normal in P, we have $P \subset N$. Since any subgroup of G containing P clearly satisfies the hypotheses of 1.5, we see that N is p-nilpotent as G is a minimal counterexample.

Now N/B is the normalizer of the central subgroup A_1/B of P/B. Moreover A_1/B is weakly closed in P/B because A_1 is weakly closed in P by hypothesis. So by the theorem of Grün the restriction is an isomorphism: $H^1(G/B) \simeq H^1(N/B)$. (In fact the restriction is an isomorphism in all dimensions [5, Appendix].) Consider the map of inflation-restriction exact sequences

$$\begin{tikzcd}
0 \arrow[r] & H^1(G/B) \arrow[r] \arrow[d, "\sim"] & H^1(G) \arrow[r] \arrow[d] & H^1(B)^G \arrow[r] \arrow[d, "\upsilon"] & H^2(G/B) \arrow[d, hook] \\
0 \arrow[r] & H^1(N/B) \arrow[r] & H^1(N) \arrow[r] & H^1(B)^N \arrow[r] & H^2(N/B)
\end{tikzcd}$$

where the arrow at the far right is injective because N/B contains the Sylow subgroup P/B. Because $H^1(B)$ is the dual of B as a vector space over $\mathbf{Z}/p\mathbf{Z}$, we see that the homomorphism υ of this diagram is an isomorphism provided every automorphism of B produced by an element of G is produced by an element of N. But by hypothesis $G/C_G(B)$ is a p-group, hence any such automorphism comes already from an element of P. Thus υ is an isomorphism, so $H^1(G) \simeq H^1(N)$ by the five lemma. Since N is p-nilpotent $H^1(N) \simeq H^1(P)$, so $H^1(G) \simeq H^1(P)$ and G is p-nilpotent by Tate's theorem. With this contradiction the proof of 1.5 is complete.

References

[1] M.F. Atiyah, Characters and cohomology of finite groups, Publ. I.H.E.S. 9 (1961).
[2] B. Huppert, Endliche Gruppe I (Springer, Berlin, 1967).
[3] D. Quillen, The spectrum of an equivariant cohomology ring, to be published.
[4] J. Tate, Nilpotent quotient groups, Topology 3 (1964) 109–111.
[5] R.G. Swan, The p-period of a finite group, Ill. J. Math. 4 (1960) 341–346.
[6] L. Evens, A generalization of the transfer map in the cohomology of groups, Trans. A.M.S. 108 (1963) 54–65.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 4 (1971) pp. 373–376

EVERY CATEGORY IS A FACTORIZATION OF A CONCRETE ONE

Luděk KUČERA
Charles University, Prague, Czechoslovakia

Received 13 May 1971

A category K is called *concrete* if there exists a faithful functor from K into the category of all sets and mappings.

We deal with concrete categories usually, but there are two important exceptions – the category of topological spaces and classes of homotopically equivalent continuous mappings and the category of small categories and classes of naturally equivalent functors, which are not concrete (see [1, 2]).

It is evident that these two categories are factorizations of concrete ones, where a factorization is defined as follows:

Definition. Let K be a category. A *congruence* $\sim$ on K is an equivalence relation on the class of all morphisms of K such that if a, b, c, d are morphisms of K, $a \sim b$, $c \sim d$ and ac is defined then bd is defined and $ac \sim bd$.

Note that if $a \sim b$ then both a and b have a common domain and a common range.

Definition. Let $\sim$ be a congruence on K. The category $K/\sim$ is defined as follows:

the class of objects of $K/\sim$ is the same as the class of objects of K,

morphisms of $K/\sim$ from X into Y are all equivalence classes (in the equivalence $\sim$) of morphisms of K from X into Y,

$\bar{a}\,\bar{b} = \overline{ab}$, where $\bar{a}$ ($\bar{b}$, $\overline{ab}$, resp.) is a morphism of $K/\sim$ containing a (b, ab, resp.).

$K/\sim$ is called a *factorization* of K.

The necessary and sufficient condition for the concreteness of a category was announced by Freyd [1]. The next theorem gives a characterization of categories which are factorizations of concrete categories. The problem was suggested by Z. Hedrlín.

Theorem. *Up to an isomorphism, every category is a factorization of a concrete one.*

The theorem will be proved in the Godel–Bernays set theory with the axiom of choice for classes.

Thus, from this point of view, the notion of a concrete category can be considered as basic notion in the category theory, because the category theory deals with concrete categories and categories derived from them in a simple way.

Proof of the theorem. Let K be a category. We can suppose that objects of K are ordinals, since we assume the axiom of choice for classes.

We shall prove that K is a factorization of a following concrete category L:

the class of objects of L is equal to the class of objects of K;

morphisms of L from m into n are all couples (f, g) such that there exists an object $C(f, g)$ of K such that $C(f, g) \leq m, n$ and $g : m \to C(f, g)$, $f : C(f, g) \to n$ are morphisms of K; in the case $m = n$ a further morphism, the formal identity 1_m, is added; the composition is defined as follows:

$$(f, g)(h, i) = \begin{cases} (f, ghi) & \text{for} \quad C(f, g) \leq C(h, i) \\ (fgh, i) & \text{for} \quad C(f, g) > C(h, i). \end{cases}$$

$$1_m (f, g) = (f, g)\, 1_n = (f, g).$$
$$1_m 1_m = 1_m .$$

It is evident that morphisms 1_m are identities of L, there is a set of morphisms of L from one object into another only and the composition of morphisms of L is a morphism of L.

The composition is associative, since

$$((f, g)(h, i))(j, k) = (f, g)((h, i)(j, k)) =$$

$$= \begin{cases} (f, ghijk) \text{ for } C(f, g) \leq C(h, i), C(j, k) \\ (fgh, ijk) \text{ for } C(h, i) < C(f, g), C(h, i) \leq C(j, k) \\ (fghij, k) \text{ for } C(j, k) < C(f, g), C(h, i), \end{cases}$$

the other cases are obvious.

Thus, we have proved that L is well defined. Now we shall show that L is concrete:

Mappings q, r from the class of morphisms of L into itself will be defined as follows:

$$q(1_m) = r(1_m) = 1_m,$$

$$q((f, g)) = (f, \mathrm{id}\, C(f, g)),$$

$$r((f, g)) = (\mathrm{id}\, C(f, g), g).$$

Note that domain A = domain $r(A)$, range A = range $q(A)$ for every morphism A of L; if AB is defined then

$$q(AB) = q(Aq(B)),$$

$$r(AB) = r(r(A)B)\,.$$

A covariant functor F from L into the category of sets and a contravariant functor G from L into the category of sets are defined by

$$F(m) = \{q(A) : A \quad \text{is a morphism of } L, \text{ range } A = m\},$$

$$G(m) = \{r(A) : A \quad \text{is a morphism of } L, \text{ domain } A = m\},$$

$$F(A)(B) = q(AB),$$

$$G(A)(B) = r(BA)\,.$$

Note that for $A \in F(m)$, $B \in G(m)$ it is $q(A) = A$, $r(B) = B$.

The functors F, G are well-defined, since $F(m)$, $G(m)$ are sets for every object m of L.

$$F(1_m)(B) = q(B) = B \quad \text{for every} \quad B \in F(m),$$

$$G(1_m)(B) = r(B) = B \quad \text{for every} \quad B \in G(m),$$

$$F(A)F(B)(C) = F(A)(q(BC)) = q(Aq(BC)) = q(ABC) = F(AB)(C),$$

$$G(A)G(B)(C) = G(A)(r(CB)) = r(r(CB)A) = r(CBA) = G(BA)(C).$$

It is evident that $F(A) = F(B)$ and $G(A) = G(B)$ implies $A = B$ for every morphisms A, B of L.

Therefore the disjoint union of F and HG, where H is a one-to-one contravariant functor from the category of sets into itself, is a faithful functor from L into the category of sets.

Finally, let $\sim$ be a congruence on L defined by

$$1_m \sim 1_n \qquad \text{if and only if } m = n$$

$$1_m \sim (f, g) \qquad \text{if and only if } fg = \mathrm{id}\, m$$

$$(f, g) \sim (h, i) \qquad \text{if and only if } fg = hi\,.$$

It is evident that $\sim$ is a congruence on L.

The isofunctor I from $L/\sim$ into K will be defined as follows:

$$I(m) = m,$$

$$I(\overline{1}_m) = \mathrm{id}\, m, \qquad I(\overline{(f, g)}) = fg\,.$$

I is an isofunctor, since if f is a morphism of K from m into n then

$$\begin{array}{llll} \text{either} & m \leqq n & \text{and} & f = I((f, \overline{\mathrm{id}\, m})) \\ \text{or} & n \leqq m & \text{and} & f = I((\overline{\mathrm{id}\, n, f})) . \end{array}$$

The details are left to the reader.

References

[1] P. Freyd, Homotopy is not concrete, Lecture Notes in Mathematics, Vol. 168, 25–34.
[2] P. Freyd, On the concreteness of certain categories, Volume for Rome conference, March 1969.

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 4 (1971) pp. 377–383

ON RELATIVE INJECTIVE ENVELOPES

Sabah FAKIR
Dalhousie University, Halifax (N.S.), Canada

Received 10 May 1971

§0. Introduction

In this paper we discuss a general theory of injective envelopes relative to a proper class of monomorphisms in an Abelian category. The first section deals with relative essential monomorphisms as defined by Maranda [4] and Stenström [8]. The second section contains a general comparison theorem which may be applied every time one has a reflective full subcategory of an Abelian category. Thus using the theorem of Popescu–Gabriel [7], we get more easily than in [5], the existence of enough injectives and of injective envelopes in a Grothendiek category, and we get further informations. In Section 3, we apply also our theorem to prove the new result of existence of pure-injective envelopes in a locally finitely presented Grothendieck category.

§1. Proper classes and relative essential monomorphisms

Let $\mathcal{A}$ be an Abelian category. Let $\mathcal{E}$ be a *proper class* of monomorphisms in $\mathcal{A}$, that is a class of monomorphisms satisfying the following conditions:
P1. Every split monomorphism is in $\mathcal{E}$.
P2. $\mathcal{E}$ is closed under composition.
P3. If $gf \in \mathcal{E}$, then $f \in \mathcal{E}$.
P4. $\mathcal{E}$ is closed under pushout, i.e. if

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[d, "k"] \\ C \arrow[r, "h"'] & D \end{tikzcd}$$

is a pushout, and if $f \in \mathcal{E}$, then $h \in \mathcal{E}$.

An element of this class is called an $\mathcal{E}$-monomorphism. An object Q of $\mathcal{A}$ is called *$\mathcal{E}$-injective*, if for every $f \in \mathcal{E}$, $f: A \to B$, and every morphism $g: A \to Q$, there exists a morphism $h: B \to Q$ such that $hf = g$. We say that $\mathcal{A}$ has *enough $\mathcal{E}$-injectives*, if for every object A, there exists $f: A \to I$, with $f \in \mathcal{E}$, and I $\mathcal{E}$-injective. When $\mathcal{E}$ = class of all monomorphisms, we say simply injective.

The following definitions are due to Maranda [4] and Stenström [8].

Definitions

1) An $\mathcal{E}$-monomorphism $f: L \to M$ is called *$\mathcal{E}$-essential*, or an $\mathcal{E}$-essential extension of L, if every morphism $g: M \to N$, such that $gf \in \mathcal{E}$, is a monomorphism.

The class of $\mathcal{E}$-essential monomorphisms is denoted $\mathcal{E}_e$.

2) Let $f: L \to M$ be an element of $\mathcal{E}_e$. f is called *maximal* if every morphism $g: M \to N$, such that $gf \in \mathcal{E}_e$, is an isomorphism. It is called an *$\mathcal{E}$-injective envelope* of L, if M is $\mathcal{E}$-injective.

We denote $\mathcal{E}_{e,m}$ the class of maximal elements of $\mathcal{E}_e$.

Remarks

1) Let $f: L \to M$ be in $\mathcal{E}$. f is in $\mathcal{E}_e$ if and only if every epimorphism $g: M \to N$, such that $gf \in \mathcal{E}$, is an isomorphism.

2) A split monomorphism which is $\mathcal{E}$-essential is an isomorphism.

We recall that a category is said to be *well-powered* if every object has only a set of subobjects. *We assume for the remaining of this section that $(\mathcal{A}, \mathcal{E})$ satisfies the following condition (*).*

Condition (*). $\mathcal{A}$ is well-powered, has exact direct colimits and the class $\mathcal{E}$ is closed under direct colimits.

Remark 3. A direct colimit in $\mathcal{A}$ is a colimit of a functor $\mathcal{D} \to \mathcal{A}$, where $\mathcal{D}$ is a small *directed category*, that is a small category such that

1) For every pair of objects D, E, there exists an object F and morphisms $D \to F$, $E \to F$.

2) For every pair of morphisms $f, g: D \to E$, there exists a morphism $h: E \to F$, such that $hf = hg$.

By [1, Exposé 1, Proposition 8.1.6] there exists then an ordered directed set $\mathcal{I}$ and a cofinal functor $\mathcal{I} \to \mathcal{D}$.

Proposition 1. *For every $f \in \mathcal{E}$, $f: L \to M$, there exists an epimorphism $g: M \to N$, such that $gf \in \mathcal{E}_e$.*

Proof. Consider the set S of (classes of) epimorphisms $g: M \to ?$, such that $gf \in \mathcal{E}$. S is not empty, because $\mathrm{id}_M \in S$. We define an order relation in S by $g \leq h$ if there exists a morphism k such that $kg = h$. S is inductive, indeed if $\{g_i \mid i \in I\}$ is a linearly

ordered subset of S, then $g = \operatorname{colim} g_i$, is an element of S, and $g \geq g_i$ for every $i \in I$. Let $m : M \to N$ be a maximal element in S; consider an epimorphism $k : N \to K$, such that $kmf \in \mathcal{E}$. Then $km \in S$, hence k is an isomorphism. This proves that $mf \in \mathcal{E}_s$.

Corollary 1. *Let $f : L \to M$ be in $\mathcal{E}_{sm}$. If $gf \in \mathcal{E}$, then g is a split monomorphism.*

Lemma 1. *A colimit of a direct system of $\mathcal{E}$-essential monomorphisms $f_i : L \to M_i$, $i \in I$, is an $\mathcal{E}$-essential monomorphism.*

Proof. Let $f : L \to M$ be a colimit of this system, and let $m_i : M_i \to M$, be the canonical morphisms. Then $f \in \mathcal{E}$. Consider a morphism $g : M \to N$ such that $gf \in \mathcal{E}$. Then $gm_if_i \in \mathcal{E}$, hence gm_i is a monomorphism for every $i \in I$. Since $g = \operatorname{colim}_i gm_i$, g is also a monomorphism.

Proposition 2. *For every object Q of $\mathscr{A}$, the following conditions are equivalent.*
a) *Q is $\mathcal{E}$-injective.*
b) *Every $f : Q \to M$, $f \in \mathcal{E}$, is split.*
c) *There exists an object L, and $f : L \to Q$, $f \in \mathcal{E}_{sm}$.*
d) *Every $g : Q \to M$, $g \in \mathcal{E}_s$, is an isomorphism.*

Proof. a) $\Rightarrow$ b). The identity of Q extends to a morphism $g : M \to Q$, such that $gf = \mathrm{id}_Q$.

b) $\Rightarrow$ c). Let $f : Q \to M$ be in $\mathcal{E}_s$, then f is a split monomorphism hence an isomorphism (Remark 2). This proves that $\mathrm{id}_Q \in \mathcal{E}_{sm}$.

c) $\Rightarrow$ d). Let $f : L \to Q$ be in $\mathcal{E}_{sm}$, and $g : Q \to M$ in $\mathcal{E}_s$. Then $gf \in \mathcal{E}$, hence by Corollary 1, g is a split monomorphism, and so an isomorphism.

d) $\Rightarrow$ a). Consider an $\mathcal{E}$-monomorphism $f : A \to B$, and a morphism $g : A \to Q$. We form a pushout

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[d, "k"] \\
Q \arrow[r, "h"] & P
\end{tikzcd}$$

$h \in \mathcal{E}$, because $\mathcal{E}$ is a proper class. By Proposition 1, there exists a morphism $p : P \to K$ such that $ph \in \mathcal{E}_s$. Hence ph is an isomorphism, and so there exists $m : P \to Q$ such that $mh = \mathrm{id}_Q$, and $mkf = g$.

Corollary 2. a) *Any maximal $\mathcal{E}$-essential monomorphism is an $\mathcal{E}$-injective envelope.*
b) *Let $f : L \to M$ and $g : L \to N$, be in $\mathcal{E}_{sm}$. There exists an isomorphism $h : M \to N$ such that $hf = g$.*

c) If L has a maximal $\mathcal{E}$-essential extension, then any $\mathcal{E}$-injective envelope of L is in $\mathcal{E}_{s,m}$.

Proposition 3. *If, furthermore, $\mathcal{A}$ has enough $\mathcal{E}$-injectives, then $\mathcal{A}$ has maximal $\mathcal{E}$-essential extensions.*

Proof. The proof follows an argument of Maranda [4, Theorem 5]. Assume that an object L has no maximal $\mathcal{E}$-essential extension. Then we construct, for every ordinal α, a morphism $f_\alpha : L \to M_\alpha$, $f_\alpha \in \mathcal{E}_s$, and for every pair α, β of ordinals, such that $\alpha < \beta$, a monomorphism $g_{\alpha,\beta} : M_\alpha \to M_\beta$, which is not an epimorphism, as follows.

1) $M_0 = L$ and $f_0 = \mathrm{id}_L$.

2) For an ordinal α, assume f_β and $g_{\beta\gamma}$ defined for all $\beta, \gamma < \alpha$. If α is not limit, we choose a monomorphism $h : M_{\alpha-1} \to A$, such that $hf_{\alpha-1} \in \mathcal{E}_s$, and that h is not an isomorphism. We set $M_\alpha = A$, $g_{\alpha-1,\alpha} = h$, and $g_{\beta\alpha} = hg_{\beta,\alpha-1}$ for $\beta < \alpha - 1$. If α is a limit ordinal, we set $f_\alpha = \operatorname{colim} f_\beta$, $\beta < \alpha$, and $g_{\beta\alpha}$ the canonical colimit morphisms; $g_{\beta\alpha}$ is a monomorphism since f_β is in $\mathcal{E}_s$, and it is not an isomorphism.

Now, by assumption, there exists $m : L \to I$, with $m \in \mathcal{E}$, and I $\mathcal{E}$-injective. Hence for each α, there exists a morphism $m_\alpha : M_\alpha \to I$, such that $m_\alpha f_\alpha = m$, and m_α is a monomorphism since $f_\alpha \in \mathcal{E}_s$.

So that one gets for every ordinal α

$$\operatorname{Card} \alpha \leq \operatorname{Card} \{\text{subobjects of } M_\alpha\} < \operatorname{Card} \{\text{subobjects of } I\},$$

but this is in contradiction with the assumption that $\mathcal{A}$ is well-powered.

Example. The category of right modules over a ring R, has injective envelopes. This result is due originally to Eckmann–Schopf [3].

§2. Theorem of comparison

Theorem 1. *Consider a functor $T : \mathcal{A} \to \mathcal{A}'$, a proper class $\mathcal{E}$ in $\mathcal{A}$ and a proper class $\mathcal{E}'$ in $\mathcal{A}'$.*

Assume that $\mathcal{A}$ and $\mathcal{A}'$ are Abelian, that T is full, faithful, has a left adjoint S, and that $T\mathcal{E} \subset \mathcal{E}'$, $S\mathcal{E}' \subset \mathcal{E}$. Then for any object L and any morphism $f : L \to M$ in $\mathcal{A}$, one has

a) *$f \in \mathcal{E}$ if and only if $Tf \in \mathcal{E}'$.*

b) *$f \in \mathcal{E}_s$ if and only if $Tf \in \mathcal{E}'_s$.*

c) *$f \in \mathcal{E}_{s,m}$ if $Tf \in \mathcal{E}'_{s,m}$.*

d) *If $(\mathcal{A}, \mathcal{E})$ satisfies the condition (*) of Section 1, and if TL has a maximal $\mathcal{E}'$-essential extension, then L has a maximal $\mathcal{E}$-essential extension, and $Tf \in \mathcal{E}'_{s,m}$ as soon as $f \in \mathcal{E}_{s,m}$.*

e) *L is $\mathcal{E}$-injective if and only if TL is $\mathcal{E}'$-injective.*

Proof. Let $\alpha : \mathrm{id}_{\mathcal{A}'} \to TS$ and $\beta : ST \to \mathrm{id}_{\mathcal{A}}$, be the adjunction morphisms. We may assume that $ST = \mathrm{id}_{\mathcal{A}}$ and $\beta = \mathrm{id}$.
a) This follows from the fact that $STf = f$.
b) Assume $f \in \mathcal{C}_s$. Let $h : TM \to N$ be a morphism such that $hTf \in \mathcal{C}'$; then $S(hTf) = Sh \circ f \in \mathcal{C}$; therefore Sh is a monomorphism, and also TSh. Since $\alpha_N h = TSh$, h is a monomorphism.

$$\begin{tikzcd}
TM \arrow[r, "h"] \arrow[d, "\mathrm{id}"'] & N \arrow[d, "\alpha_N"] \\
TSTM \arrow[r, "TSh"] & TSN
\end{tikzcd}$$

The converse follows from the fact that T reflects monomorphisms.
c) This follows from b) and the fact that T reflects isomorphisms.
d) Let $g : TL \to N$ be in $\mathcal{C}'_{s,m}$. Then $Sg \in \mathcal{C}$; hence by Proposition 1, there exists a morphism $h : SN \to P$, such that $k = hSg \in \mathcal{C}_s$. By b), $Tk \in \mathcal{C}'_s$. Since $Tk = Th \circ \alpha_N \circ g$, one has $Th \circ \alpha_N$ is an isomorphism. Therefore $Tk \in \mathcal{C}'_{s,m}$ and $k \in \mathcal{C}_{s,m}$. The remaining of the statement follows from Corollary 2, b).

$$\begin{tikzcd}
L \arrow[r, "sg"] \arrow[dr, "k"'] & SN \arrow[d, "h"] \\
 & P
\end{tikzcd}
\qquad
\begin{tikzcd}
TL \arrow[r, "g"] \arrow[ddr, "Tk"'] & N \arrow[d, "\alpha_N"] \\
 & TSN \arrow[d, "Th"] \\
 & TP
\end{tikzcd}$$

e) The proof is straightforward.

We recall that a family of objects U_i, $i \in I$, is said to be a *generating family* for a category $\mathcal{A}$, if for every pair of distinct morphisms $f, g : A \to B$, there exists $i \in I$ and a morphism $h : U_i \to A$ such that $fh \neq gh$.

Corollary 3. *Every Abelian category $\mathcal{A}$, with a generator U, and exact direct colimits, has enough injectives and has injective envelopes.*

Proof. $\mathcal{A}$ is well-powered by [6, Prop. 15.1, page 71], and satisfies the condition (*). Let R be the ring of endomorphisms of U, and let Mod R be the category of right R-modules. Consider the functor $T : \mathcal{A} \to \mathrm{Mod}\, R$, $TX = \mathcal{A}(U, X)$. By [7], this functor is full, faithful, and has an exact left adjoint. Since Mod R has maximal essential extensions, the corollary follows by Theorem 1,d). Moreover, T preserves and

reflects injective envelopes. This fact has also been observed by Popescu [11, Corollary 6.32].

§3. Existence of pure injective envelopes

We recall some definitions. In an Abelian category $\mathcal{A}$, an object M is called *finitely presented* if the functor $\mathcal{A}(M, -): \mathcal{A} \to \mathcal{A}b$, preserves direct colimits. Here $\mathcal{A}b$ denotes the category of Abelian groups. A short exact sequence is called *pure* if every finitely presented object is relatively projective for it; the corresponding monomorphism to this sequence is also called pure. A category is called *locally finitely presented Grothendieck category* (in short l.f.p.g.), if it is Abelian, with exact direct colimits and a generating family of finitely presented objects. Any functor category $\hat{\mathcal{X}} = (\mathcal{X}^*, \mathcal{A}b)$, with $\mathcal{X}$ small, is l.f.p.g.

Throughout this section, *$\mathcal{A}$ denotes a l.f.p.g. category*; $\mathcal{P}$ denotes the full subcategory of $\mathcal{A}$ whose objects are the finitely presented objects of $\mathcal{A}$.

Proposition 4.
1) *$\mathcal{P}$ is additive, has cokernels and is equivalent to a small category.*
2) *For every object A in $\mathcal{A}$, the category $\mathcal{P}/A$ is directed and A is a canonical colimit of the functor $V: \mathcal{P}/A \to \mathcal{A}$, $V(f: P \to A) = P$.*
3) *The functor $T: \mathcal{A} \to \hat{\mathcal{P}}$, $TX = \mathcal{A}(-, X)$, is full, faithful and has an exact left adjoint.*

Proof. It is done in [2]. To construct the left adjoint S, one remarks that every F in $\hat{\mathcal{P}}$ is a colimit of the functor $L: \mathcal{P}/F \to \hat{\mathcal{P}}$, $L(\mathcal{P}(-, P) \to F) = \mathcal{P}(-, P)$, and one sets SF = colimit of the functor $V: \mathcal{P}/F \to \mathcal{A}$, $V(\mathcal{P}(-, P) \to F) = P$.

Remark. We shall not use the exactness of S.

The following lemmas are stated in [9], and are easy consequences of the definitions, and of Proposition 4, 2).

Lemma 2. *The pure monomorphisms of $\mathcal{A}$ form a proper class $\mathcal{E}$ and $(\mathcal{A}, \mathcal{E})$ satisfies the condition $(*)$ of Section 1.*

Lemma 3. *A monomorphism in $\mathcal{A}$ is pure if and only if it is a colimit of a direct system of split monomorphisms.*

Theorem 2. *$\mathcal{A}$ has enough pure-injectives and has pure-injective envelopes.*

Proof. The functor $T: \mathcal{A} \to \hat{\mathcal{P}}$, considered in Proposition 4, preserves direct colimits; in fact if $B_i \to B$, $i \in I$, is a direct colimit system, the canonical morphism

Colim $TB_i \to TB$, is an isomorphism, since its value for every object P of $\mathcal{P}$ is an isomorphism. It follows from Lemma 3, that T and its left adjoint S take pure monomorphisms into pure monomorphisms. Since $\hat{\mathcal{P}}$ has pure-injective envelopes [9, Corollary of Theorem 2], which are maximal pure-essential extensions (Prop. 3 and Corollary 2, c)), the result follows by Theorem 1, d).

Example. Let E be a finitely generated module over a commutative Noetherian ring A. We define a topology on E by taking as neighbourhoods of 0 the submodules IE, where I is a finite intersection of powers of the maximal ideals of the ring. The completion of E is then a pure-injective envelope of E in Mod A.

For this result and for a study of pure-injectives in categories of modules, we refer the reader to [10].

The author acknowledges valuable discussions with J. Giraud.

References

[1] M. Artin, A. Grothendieck and J.L. Verdier, Séminaire de Géométrie Algébrique de l'I.H.E.S. 1963–1964.

[2] S. Breitsprecher, Local endliche präsentierbare Grothendieck-Kategorien, Mitt. Math. Seminar Giessen 85 (1970) 1–25.

[3] B. Eckmann and A. Schopf, Über injective Moduln, Arch. Math. 4 (1953) 75–78.

[4] J.M. Maranda, On pure subgroups of Abelian groups, Arch. Math. 11 (1960) 1–13.

[5] B. Mitchell, The full imbedding theorem, Amer. J. Math. 86 (1964) 619–637.

[6] B. Mitchell, Theory of categories (Academic Press, New York, 1965).

[7] N. Popescu and P. Gabriel, Caractérisation des catégories abéliennes avec générateurs et limites inductives exactes, C.R. Acad. Sc. Paris 258 (1964) 4188–4190.

[8] B. Stenstrom, Pure submodules, Arkiv. Mat. 7 (1967) 159–171.

[9] B. Stenstrom, Purity in functor categories, J. Algebra 8 (1968) 352–361.

[10] R.B. Warfield Jr., Purity and algebraic compactness for modules, Pacific J. Math. 28 (1969) 699–719.

[11] I. Bucur and A. Deleanu, Introduction to the theory of categories and functors (Wiley, New York, 1968).

JOURNAL OF PURE AND APPLIED ALGEBRA – Volume 1, No. 4 (1971) pp. 385 - 391

COPRODUCTS AND A_n-CLASSES OF CERTAIN FIBER SPACES *

Chao-kun CHENG
SUCP Postdam, N.Y., U.S.A.

Received 24 May 1971

In this note we shall study the coproduct of some elements in $H^*(\Omega E; Z_p)$ for the principal fibration $\Omega K \to E \to B$ which is a pull back of a path fibration via a map f as indicated in the following diagram, where K is an Eilenberg–Maclane space:

$$\begin{tikzcd}
\Omega E \arrow[r] \arrow[d, "i"'] & \Omega K \arrow[d] \\
E \arrow[r] \arrow[d, "P"'] & PK \arrow[d] \\
B \arrow[r, "f"] & K
\end{tikzcd}$$

We require B and ΩK both to be simply connected and $H^*(B; Z_p)$ to be of finite type. Let p be a fixed prime throughout this note, and all cohomology algebras and any other algebraic gadgets are with coefficients Z_p.

In the special case in which the fibration is a stable 2-stage Postnikov system (i.e. B and K are generalized Eilenberg–Maclane spaces, and f is primitive) the papers of Massey and Peterson [6] and Smith [10] show, in order to study the Hopf algebra structure of $H^*(\Omega E)$, it is sufficient to determine coproducts on certain elements of $H^*(\Omega E)$. Our results will be sufficient enough to tell what coproducts of those elements are. Various results have been done for this special case, such as for $p = 2$ [1, 3 and 5], for p is odd [4].

The main tool we use here differs from those in [3, 4, 5]; namely we use the spectral sequence described in [2, 8 and 10]. Therefore in working out the determination of coproduct we can check the A_n-classes as well (c.f. [11] for definition of A_n-class).

* Based on a thesis submitted in partial fulfillment of the requirements for the Ph.D. Degree at the University of Notre Dame under Professor J.D. Stasheff.

§ 1. Main result

Milgram [7] and Stasheff [10] have proved that $\bar{B}C_*(\Omega X) \approx C_*(X)$ as chain complex where $\bar{B}$ is the bar construction. The differential in $\bar{B}$ is equal to $d_E + d_I$ where

$$d_E([a_1|...|a_n]) = \sum_{i=1}^{n-1} (-1)^{s(i)}[a_1|...|a_i a_{i+1}|...|a_n]$$

and

$$d_I([a_1|...|a_n]) = \sum_{i=1}^{n} (-1)^{s(i-1)}[a_1|...|da_i|...|a_n]$$

in which $s(i) = \deg[a_1|...|a_i]$ and d is the differential in $C_*(\Omega X)$.

We know from [7] that there exists a diagonal approximation of X which corresponds to the obvious diagonal map in $\bar{B}C_*(\Omega X)$. Moreover Moore [8] shows that if $H_*(\Omega X)$ is of finite type and if $\bar{B}C_*(\Omega X)$ is filtered by bars, we obtain a homology spectral sequence $E^*(\Omega X)$ which has $E^1 \approx \bar{B}(H_*(\Omega X))$ with d^1 the loop multiplication and $E^2 \approx \mathrm{Tor}^{H_*(\Omega X)}(Z_p, Z_p)$ and $E^r \to E^\infty \approx E^0(H_*(X))$ as coalgebras, where E^0 denotes the associated graded coalgebra under some filtration. Dually speaking there exists a cohomology spectral sequence $E_*(\Omega X)$ such that $E_1 \approx \bar{B}^*(H^*(\Omega X))$ and $m^*(x) = 1 \otimes x + x \otimes 1 + d_I x$ where $x \in H^*(\Omega X)$, m is the loop multiplication on ΩX. Further $E_2 \approx \mathrm{Ext}_{H_*(\Omega X)}(Z_p, Z_p)$ and $E_r \to E_0(H^*(X))$ as algebras.

Let F_r, $0 \leq r$, denote the filtration on $H^*(B)$ such that $E_r(\Omega B)$ converges to $E_0(H^*(B))$. Therefore $F_r \supset F_{r'}$ if $r' > r$. Since $H^*(B)$ is of finite type, we can choose a minimal ideal base $X = \{x_1, x_2, ...\}$ for $H^*(B)$. This base X is minimal in the sense that $\overline{H^*(B)} \neq (x_1, ..., \underline{x_t}, ...)$ for any t, such that the projection of $X \cap F_{r-1}$ into $[F_{r-1}, \overline{H^*(B)}^2]/[F_r, \overline{H^*(B)}^2]$ form a base for $[F_{r-1}\ \overline{H^*(B)}]/[F_r, \overline{H^*(B)}]$ where $[F_r, \overline{H^*(B)}^2]$ denotes the subspace of $H^*(B)$ generated by F_r and $\overline{H^*(B)}^2$.

Definition 1.1. For any $x = x_1 x_2 ... x_n$ where $x_i \in X$, $1 \leq i \leq n$, we define the weighted length of x to be $\Sigma_{i=1}^n r_i$ where $x_i \in F_{r_i}$ but $x_i \notin F_{r_i+1}$. The weighted length of $\Sigma x_1 x_2 ... x_n$ is defined to be the minimum of the weighted lengths of each term in the summation.

We use $\mathrm{wl}(x)$ to denote the weighted length of x.

As usual σ and τ are the suspension and transgression of a path fibration respectively and m denotes the loop multiplication on ΩE.

In the following theorem, let $\Omega K \to E \to B \to K$ be a principal fibration as before. For any $u \in H^*(\Omega K)$, $u = \alpha\,\iota_{\Omega K} \in H^*(\Omega K)$ where $\iota_{\Omega K}$ is the fundamental class of $H^*(\Omega K)$, and $\alpha \in \mathfrak{A}(p)$, the Steenrod algebra.

Let $u^* = \alpha\,\iota_K$ and $u' = f^*(\alpha\,\iota_K)$.

Theorem 1.2. *Consider any $v \in H^*(\Omega E)$ such that $(\Omega i)^* v = \sigma u$ for some $u \in H^*(\Omega K)$ with non-trivial transgression in the fibration $\Omega K \to E \to B \to K$. Express u' as $\Sigma^h a_i z_i + \Sigma b_j w_j$ where z_i and w_j are non-zero elements in $\bigcup_{n=1} X^n$ such that $\mathrm{wl}(z_i) = \mathrm{wl}(u')$ and $\mathrm{wl}(w_j) > \mathrm{wl}(u')$, a_i and b_j are non-zero elements in Z_p for all i and j. We obtain the following results.*

(i) *If $\mathrm{wl}(u') > 2$, there exists $w \in PH^*(\Omega E)$ such that $(\Omega i)^* w = \sigma u$.*

(ii) *If $\mathrm{wl}(u') = 2$, $z_i = x_i y_i$ where $1 \leq i \leq t$ and $x_i, y_i \in X$, and $z_i \in X$ for $t < i \leq h$. Then there exists $w \in H^*(\Omega E)$ such that $(\Omega i)^* w = \sigma u$ and $m^*(w) = 1 \otimes w + w \otimes 1 + \Sigma^t a_i (\Omega p)^* (\sigma x_i \otimes \sigma y_i) + \Sigma_{i=t+1}^h a_i (\Omega p)^* (u_i \otimes v_i)$* for some $u_i, v_i \in H^*(\Omega B)$.

(iii) *w in* (i) *and* (ii) *is an A_t-class for some $t \geq \mathrm{wl}(u') - 1$.*

§ 2. Proof of Theorem 1.2.

First we shall work on a special case; namely the principal fibration $\Omega K \xrightarrow{i} E \xrightarrow{p} B \xrightarrow{f} K$ with $K = K(Z_p, n)$ and $\Omega f \simeq *$. Then we will use the results of this special case to prove the theorem.

Lemma 2.1. *Let A, C be DGA associative connected algebra of finite type. Suppose there exists a surjective DGA algebra map $\pi : C \to A$ such that π is an isomorphim when restricted to $C_i \to A_i$ when $i < n$ and $\dim(\ker(\pi : C \to A_n)) = 1$ and the generator of $\ker(\pi : C_n \to A_n)$ is a cycle. (C_i, A_i mean the homogeneous subspace of degree i of C and A respectively.) Then in the cohomology spectral sequence (obtained from $\bar{B}(C)$ and $\bar{B}(A)$ filtered by bars) we have*

(i) *For all r, $\pi^* : E_r^{s,t}(A) \to E_r^{s,t}(C)$ is an isomorphism if $s + t < n + 1$ or $s + t = n + 1$ and $s > 1$.*

(ii) *For all r, $\pi^* : E_r^{s,n+2-s}(A) \to E_r^{s,n+2-s}(C)$ is an isomorphism when $s > 2$ except possible for one and only one s and $r > s$. Moreover in this possible case π^* is onto and $\pi^* : E_r^{1,n}(A) \to E_r^{1,n}(C)$ is an isomorphism when $r > s$ and $d_{s-1}(x) \neq 0$ where x is the class in $E_{s-1}^{1,n}(C)$ containing the dual of generator of the kernel $(\pi : C_n \to A_n)$. In other words in this possible case for any $y \in E_{s-1}^{s,n+2-s}(A)$ such that y [illegible]s to E_s term and $\pi^*(y)$ vanishes in $E_s(C)$, then there is $z \in E_{s-1}^{1,n}(C)$ such that $d_{s-1}(z) = \pi^*(y)$ and $z \notin \pi^*(E_{s-1}^{1,n}(A))$.*

Proof. The proof follows from routine work on spectral sequence.

Let $\Omega K \xrightarrow{i} E \xrightarrow{p} B \xrightarrow{f} K = K(Z_p, n)$ be a principal fibration with $\Omega f \simeq *$ and let X be a minimal ideal base for $H^*(B)$ which is chosen as in § 1. As usual ι_n will be the fundamental class in $H^*(Z_p, n; Z_p)$.

Lemma 2.2. *Assume $f^*(\iota_n) = \Sigma^h a_i z_i + \Sigma b_j w_j$ where z_i, w_j are non-zero elements in $\bigcup_{n=1} X^n$, and $s = \mathrm{wl}(z_i) = \mathrm{wl}(f^* \iota_n)$, $\mathrm{wl}(w_j) > s$, and $a_i b_j$ are non-zero elements in Z_p for all i and j. Then*

(i) $p^* f^* \iota_n = 0$.

(ii) *Consider any* $x = x_1 x_2 \dots x_l y_1 y_2 \dots y_t \in H^*(B)$ *where* $x_i, y_j \in X$ *but* $x_i \notin F_2$ *and* $y_j \in F_2$. *Each* y_j *can be represented by some* $\xi_j \in E_1^{r_j,*}(\Omega B)$ *where* $r_j = \mathrm{wl}(y_j)$. *Then* x *can be represented by* $[\sigma x_1 | \dots | \sigma x_l] \otimes \otimes \xi_j$ *in* $E_1^{\mathrm{wl}(x),*}(\Omega B)$.

Let $[\sigma x_1 | \dots | \sigma x_l] \otimes \otimes \xi_j$ *be denoted by* $x!$.

(iii) *Note* $\Omega E \simeq B \times \Omega K$ *and has loop multiplication. In* $E_*(\Omega E)$, *we have* $d_{s-1}[1 \otimes \iota_{n-2} + \mathrm{Junk}] = \lambda \Sigma a_i (\Omega p)^* z_i!$ *where* $0 \neq \lambda \in Z_p$ *and* Junk $\in \mathrm{Im}(\Omega p^*)$.

Proof. (i) Trivial.

(ii) Since $x_i \in X$ but $x_i \notin F_2$, $[\sigma x_i]$ represents x_i in $E_1^{1,*}(\Omega B)$ for all i, and since $\mathrm{wl}\, y_j = r_j$, y_j is represented by some element $\xi_j \in E_1^{r_j,*}(\Omega B)$. But at each stage E_r is a DGA algebra and $E_{r+1} = H(E_r)$. Therefore the product in E_{r+1} can be represented by the product of the representatives in E_r. Moreover E_r converges to $E_0(H^*(B))$, and hence x can be represented by $[\sigma x_1 | \dots | \sigma x_l] \otimes \otimes \xi_j$.

(iii) By looking at the filtration on the cohomology spectral sequence ΩB and (ii), we know $f^*(\iota_n)$ can be represented by $\Sigma^h a_i z_i!$ in $E_1^{s,n-s}(\Omega B)$. Since $p^* f^*(\iota_n) = 0$, it follows that in the cohomology spectral sequence of ΩE, $(\Omega p)^*(\Sigma a_i z_i!)$ vanishes in certain stage. But $\Sigma a_i z_i!$ survivies to $E_\infty(\Omega B)$; hence the map induced by Ωp from $E_r^{s,n-s}(\Omega B)$ to $E_r^{s,n-s}(\Omega E)$ can not be one-to-one when $r \geq s$. From Lemma 2.1, there exists some element $[z] \in E_1^{1,n-2}$ such that $z \in H^*(\Omega E)$, $z \notin \mathrm{Im}(\Omega p)^*$ and $d_{s-1}[z] = \lambda (\Omega p)^* (\Sigma a_i z_i!)$ for some $0 \neq \lambda \in Z_p$. Of course such z can be written as $1 \otimes \iota_{n-2}$ + Junk with suitable λ.

Proof of Theorem 1.2. Consider $v \in H^*(\Omega E)$ such that $(\Omega i)^* v = \sigma u$ for some $u \in H^*(\Omega K)$ and u does not have trivial transgression in the fibration $\Omega K \xrightarrow{i} E \xrightarrow{p} B \xrightarrow{f} K$ and $u' = \Sigma^h a_i z_i + \Sigma b_j w_j$ as before. By the definition of u' there is a $u^* \in H^*(K)$ such that $f^* u^* = u'$. Consider the following diagram:

$$\begin{tikzcd}
\Omega K \arrow[r, "i"] \arrow[d, dashed, "u"'] & E \arrow[r, "p"] \arrow[d, dashed, "g"] & B \arrow[r, "f"] \arrow[d, dashed, "\mathrm{id}"] & K \arrow[d, dashed, "u^*"] \\
K(Z_p, \deg u) \arrow[r] & \bar{E} \arrow[r, "\bar{p}"] & B \arrow[r, "u'"'] & K(Z_p, 1 + \deg u).
\end{tikzcd}$$

Since $f^* u^* = u'$, therefore we can construct a fibration map g such that $g | \Omega K = u$. Looping the diagram we get another pair of principal fibrations with multiplicative fiber map Ωg. Since $v \in H^*(\Omega E)$ therefore in $\Omega^2 K \xrightarrow{\Omega i} \Omega E \xrightarrow{\Omega p} \Omega B \xrightarrow{\Omega f} \Omega K$, σu transgresses to zero, i.e. $\Omega f^* u = 0$. Therefore $\sigma u' = \sigma(f^* u^*) = (\Omega f)^*(\sigma u^*) = (\Omega f)^* u = 0$. So the bottom fibration is of the type we discussed in Lemma 2.2. Since Ωg is a multiplicative map, it preserves the cohomology spectral sequence. Hence we shall apply Lemma 2.2 on the bottom fibration and then carry the properties to ΩE via $(\Omega g)^*$.

Let $w = (\Omega g)^*(\iota_{\deg u-1} + \text{Junk})$ where Junk $\in \text{Im}(\Omega p)^*$. It follows that $(\Omega i)^* w = \sigma u$. Then in $E_*(\Omega E)$ we have

a) $d_r[w] = 0$ if $r < s-1$,

b) $d_{s-1}[w] = d_{s-1}((\Omega g)^*[\iota_{\deg u-1} + \text{Junk}]) = (\Omega g)^*(\lambda(\Omega p)^* \Sigma a_i z_i!) = \lambda \Sigma a_i(\Omega p)^* z_i!$ for some non-zero λ.

Therefore w is an A_t-class for some $t \geq s-1$, so we have established (iii) in the theorem.

Since $d_1 x = m^* x - (1 \otimes x + x \otimes 1)$ for any $x \in H^*(\Omega E)$, (i) and (ii) in the theorem follows immediately.

§ 3. Stable two stage Postnikov system

In this section B is a generalized Eilenberg -Maclane space and f is an H-map. We can choose the obvious minimal ideal base X for $\overline{H^*(B)}$ by taking X to be the collection of the elements of form $\alpha\iota$ where α is an admissible element in $\mathfrak{A}(p)$ with excess less than dim ι and ι the fundamental classes of $H^*(B)$. For stable two-stage Postnikov system we will have a stronger theorem than Theorem 1.2 as follows:

Theorem 3.1. *In Theorem 1.2, if the fibration is a stable two-stage Postnikov system then in* (iii) *we have*

(a) *With Z_2 as coefficient group, then $w \in H^*(\Omega E)$ as found in* (i) *or* (ii) *is an A_t-class and not an A_{t+1}-class where $t = \text{wl}(u') - 1$.*

(b) *With Z_p, p odd, as coefficient group, and assuming $\sigma(z_{i,m}) \neq 0$ for all i, m where $u' = \Sigma a_i z_i + \Sigma b_j w_j$ and $z_i = z_{i,1} z_{i,2} \cdots z_{i,g_i}$ and $z_{i,m} \in X$, $w_j \in \bigcup_{n=1} X^n$, then $w \in H^*(\Omega E)$ as found in* (i) *or* (ii) *is an A_t-class but not an A_{t+1}-class where $t = \text{wl}(u') - 1$.*

The reason why we can not have this stronger version in general case comes from the fact that $(\Omega g)^* \Sigma \Omega p^* a_i z_i!$ may equal zero, or may get killed before t-stage. So we shall prove those are not the case when the fibration is a two-stage Postnikov system. Before proceeding to the proof, let us recall some results about the cohomology spectral sequence. We consider the dual statement of Corollary 3.5 of [2] and some statement in the proof of Theorem 4.1 or [2].

Theorem 3.2 (Clark [2]). *If X is a simply connected H-space and $H^*(X)$ is of finite type then the cohomology spectral sequence $E_*(X)$ has the following properties:*

$E_1(\Omega X) \simeq B(H^*(\Omega X))$,

$E_2(\Omega X) \simeq \text{Ext}_{H_*(\Omega X)}(Z_p, Z_p)$ *as Hopf algebras,*

$E_r(\Omega X) \to E_0(H^*(X))$ *as algebras.*

Let us take a close look at the E_2-term. $E_2(\Omega X) \simeq \text{Ext}_{\otimes A_i}(Z_p, Z_p)$ where $H_*(\Omega X) \cong \otimes A_i$ as algebras and A_i is a one generator Hopf algebra.

A	$\mathrm{Ext}_A(Z_p, Z_p)$
$E(x, m)$	$L(\{x^*\}, 1, m)$
$L(x, 2m)$	$E(\{x^*\}, 1, 2m)$
$L(x, 2m)/(x^{p^k})$	$E(\{x^*\}, 1, 2m) \otimes L(tx^*, 2, 2mp^k)$

where E and L are exterior algebra and polynomial algebra respectively, and x^* is the dual of x and tx^* is the dual of the transpotence of x.

In the right column bidegrees are specified.

Theorem 3.3 (Clark [2]). *In $E_r(\Omega X)$ for $r \geq 2$ we have*

(i) *Elements of filter degree 1 are primitive and indecomposable;*

(ii) *tx^* is indecomposable;*

(iii) *If $(x^*)^k \neq 0$ but $(x^*)^{k+1} = 0$, then $k < r$;*

(iv) *As a differential Hopf algebra, E_r is the tensor product of differential Hopf algebra with differential identically zero, and differential Hopf algebra of the form $E(y, 1, m) \otimes L(z, u, v)$ where $d_r(y) = z^{p^g}$, with $r = up^g - 1$ for $u = 1$ or $u = 2$.*

Proof of Theorem 3.1. Let $v \in H^*(\Omega E)$ such that $(\Omega i)^* v = \sigma u$ for some $u \in H^*(\Omega K)$ with non-trivial transgression. Because B is a generalized Eilenberg–Maclane space, $\sigma u' = 0$ and f is an H-map, therefore u' must be

(i) $(z)^{p^g}$ for some $z \in H^*(B)$ and $g > 0$ and z is indecomposable, or

(ii) $\beta(z)$ for some $z \in H^*(B)$ such that $\sigma(z) = (w)^{p^g}$ for some $w \in H^*(\Omega E)$ and $g > 0$ when p is odd prime. It is obvious that $\sigma(\beta(z)) = 0$ in (ii), so in order to prove Theorem 3.1 we need only to examine (i).

We claim $(\Omega p)^*(\sigma z) \neq 0$. In a stable two-stage Postnikov system we have $\ker(\Omega p)^*$ = the ideal generated by the image of $(\Omega f)^*$ on indecomposable elements of $H^*(\Omega K)$ [9]. So if $(\Omega p)^*(\sigma z) = 0$ then in the Serre spectral sequence of $\Omega^2 K \to \Omega E \xrightarrow{\Omega p} \Omega B \xrightarrow{\Omega f} \Omega K$ some element in $H^*(\Omega^2 K)$ transgresses to σz. Hence in the Serre spectral sequence of $\Omega K \to E \xrightarrow{p} B \xrightarrow{f} K$ some element $x \in H^*(\Omega K)$ transgresses to z. Then it is obvious $x \otimes z^{p^g - 1}$ survives till deg z stage and $x \otimes z^{p^g - 1}$ hits z^{p^g} in this stage. We have a contradiction to $u' = z^{p^g}$.

Therefore looking back to the proof of Theorem 1.2:

$$(\Omega g)^* \, \Omega p^* [\underbrace{\sigma z | \ldots | \sigma z}_{p^g \text{ terms}}] \neq 0.$$

Now we claim this class survives till $E_{p^g - 1}$ stage. If not, by Theorem 3.3: There is an element of filter degree 1 which will hit $(\Omega p)^*[\underbrace{\sigma z | \ldots | \sigma z}_{p^h \text{ terms}}]$ with $h < g$. Since we

know any element in filter degree 1 is primitive and indecomposable. $H^*(\Omega E)$ is generated by elements in $\mathrm{Im}(\Omega p)^*$ or elements with non-zero $(\Omega i)^*$ image. So there exists some $b \in H^*(\Omega E)$ such that $(\Omega i)^*(b) = \sigma a$ for some $a \in H^*(\Omega K)$ such that $d_{p^h-1}[b] = (\Omega p)^*[\sigma z| \ldots |\sigma z]$. But in this case a has non-trivial transgression in $\Omega K \to E \to B \to K$. By the same argument as in Theorem 1.2. there exists $e \in H^*(\Omega E)$ such that $d_r[e] = a'!$ for some r and $(\Omega i)^*(e) = \sigma a$. But $e-b$ is of filter degree 1 and $(\Omega i)^*(e-b) = 0$. Hence $e-b \in \mathrm{Im}(\Omega p)^*$. Since Ωp induces a cohomology spectral sequence map from $E_r(\Omega B) \to E_r(\Omega E)$, and $e-b$ survives to $E_\infty(\Omega B)$, $e-b$ survives to $E_\infty(\Omega E)$. It follows that $r = p^h-1$ and $d_r[e] = a'! = (\Omega p)^*[\sigma z| \ldots |\sigma z] = d_r[b]$ that is $a' = z^{p^h}$. But the above will imply that in the $p^h \deg z$ stage of the Serre spectral sequence of $\Omega K \to E \to B \to K$, $a \otimes z^{p^g-p^h} \neq 0$ and hits z^{p^g} which contradicts that u trasngresses to z^{p^g} in the Serre spectral sequence of $\Omega K \to E \to B \to K$.

Therefore $(\Omega g)^*(\Omega p)^* \underbrace{[\sigma z| \ldots |\sigma z]}_{p^g \text{ terms}} = (\Omega p)^* \underbrace{[\sigma z| \ldots |\sigma z]}_{p^g \text{ terms}}$ survives to E_{p^g-1} stage and there exists some $w \in H^*(\Omega E)$ such that $d_{p^g-1}[w] = (\Omega p)^*[\sigma z| \ldots |\sigma z]$ and $(\Omega i)^* w = \sigma u$.

References

[1] C.K. Cheng, Thesis at University of Notre Dame.

[2] A. Clark, Homotopy commutativity and the Moore spectral sequence, Pacific J. Math 15 (1965) 65–74.

[3] J. Harper, On the cohomology of stable 2-stage Postnikov systems, Bull. Am. Math. Soc. 76 (1970) 807–809.

[4] J. Harper and C. Schochet, Coalgebra extensions in 2-stage Postnikov systems, Matematisk Institut, Aarhus Univ. 1970/1971.

[5] L. Kristensen, On the cohomology of spaces with two non-vanishing homotopy groups, Math. Scand. 12 (1963) 83–105.

[6] W.S. Massey and F.P. Peterson, The mod 2 cohomology structure of certain fibre spaces, Mem. Am. Soc. 74 (1967).

[7] J.R. Milgram, The bar construction and Abelian H-spaces, Ill. J. Math 11 (1967) 242–250.

[8] J.C. Moore, Seminaire H. Cartan 1959/1960.

[9] L. Smith, Cohomology on stable 2-stage Postnikov systems, Ill. J. Math 11 (1967) 310–329.

[10] J.D. Stasheff, Homotopy associativity of H-spaces (II), Trans. Am. Math Soc. (1963) 293–312.

AUTHOR INDEX